Structural and Biochemical Investigation of Methylation and Elucidation of *t*-Butyl Formation in Polyketide Biosynthesis

by

Meredith Anne Skiba

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biological Chemistry)
in the University of Michigan
2018

Doctoral Committee:

Professor Janet L. Smith, Chair
Professor Philip C. Andrews
Assistant Professor Uhn-Soo Cho
Associate Professor Patrick O'Brien
Professor David H. Sherman

Meredith A. Skiba

skibam@umich.edu

ORCID iD: 0000-0003-4615-6775

Acknowledgements

First, I would like to thank my advisor Dr. Janet L. Smith for her encouragement and guidance during my graduate career. It has truly been a privilege to be a member of her lab and learn from her. I would also like to thank to all the past and present members of the Smith lab who provided guidance and camaraderie along the way. I am especially grateful to Andy Sikkema, who gifted me with a great deal of preliminary data that inspired much of this thesis. Much of my work was assisted by a talented team of undergraduates, Marissa Bivins, Rebecca Sturgis, and Collin Tran, who I must thank for their dedication and patience, as I learned to be an effective teacher.

This thesis could not have been completed without the aid of wonderful collaborators. Nathan Moss, Dr. Lena Gerwick, and Dr. Bill Gerwick (Scripps Institution of Oceanography, University of California San Diego) were essential to the success of the apratoxin projects. Dr. Andrew Lowell of Dr. David Sherman's lab (University of Michigan), Dr. William Fiers of Dr. Courtney Aldrich's lab (University of Minnesota), and members of Dr. Peter Wipf's lab (University of Pittsburg) contributed key synthetic reagents to this thesis and provided essential chemical insight. Dr. Min Su (University of Michigan) was vital to the success of the electron miscroscopy work and was always enthusiastic in sharing his expertise. Finally, I am indebted to Dr. Wendy Feng (University of Michigan) and the staff at the GM/CA beamline at the Advanced Photon Source (Argonne National Lab) for their technical support that made this thesis possible.

I am very appreciative of the guidance from my current and past committee members: Dr. Philip Andrews, Dr. Uhn-Soo Cho, Dr. Patrick O'Brien, and Dr. David Sherman, as well as Dr. John Tesmer (past member) and Dr. Georgios Skiniotis (past member). I would also like to thank Mindy Mackey and Beth Goodwin for their administrative support while at Michigan.

Finally, I am grateful for the various funding sources that have supported my graduate research: The University of Michigan Life Sciences Institute, Department of Biological Chemistry, and Rackham Graduate School, and the NIH Cellular Biotechnology Training Program (GM008353).

Table of Contents

Acknowledgements ..... **ii**
List of Tables ..... **viii**
List of Figures ..... **x**
Abbreviations ..... **xv**
Abstract ..... **xviii**
Chapter 1 Introduction **1**
Natural Products **1**
Biosynthetic Gene Clusters **1**
Type I Polyketide Synthases **2**
PKS pathway initiation 5
Polyketide chain extension and modification 6
PKS Module Architecture and Relationship to Mammalian Fatty Acid Synthase **8**
Methyltransferases in Natural Product Biosynthesis **10**
PKS *C*-MT domains 12
PKS *O*-MT domains 13
Initiation module MTs 13
Thesis Overview **14**
Chapter 2 Domain Organization and Active Site Architecture of a Polyketide Synthase *C*-Methyltransferase **17**
Summary **17**
Introduction **17**
Experimental Procedures **20**
Construct design 20
Protein expression and purification 20
Differential scanning calorimetry (DSC) 21
Protein crystallization 21
Data collection and structure determination 22
Enzyme assays 27

Results and Discussion **28**
Identification of *C*-MT domain boundaries 28
Methyltransferase activity 31
Structure of CurJ *C*-MT 35
Active site architecture 39
Activity of CurJ *C*-MT with active site substitutions 42
Substrate modeling 42
Similarities of CurJ *C*-MT to other PKS *C*-MTs 43
Chapter 3 Structural Basis of Polyketide Synthase *O*-Methylation **44**
Summary **44**
Introduction **44**
Experimental Procedures **46**
Construct design 46
Protein expression and purification 49
Production of acyl-ACPs 50
Enzyme assays 50
Protein crystallization and structure determination 51
Results and Discussion **62**
Determining *O*-MT domain boundaries 62
O-MT activity and substrate specificity 62
Structural characterization of StiD and StiE *O*-MT 66
Probing *O*-MT active site architecture 71
Catalytic strategy of PKS *O*-MT domains 74
O-MTs in the context of a PKS module 75
Chapter 4 A Mononuclear Iron-Dependent Methyltransferase Catalyzes Initial Steps in Assembly of the Apratoxin A Polyketide Starter Unit **76**
Summary **76**
Introduction **76**
Experimental Procedures **79**
Construct design 79
Protein expression and purification 82
Production of acyl-ACPs 83
Enzyme assays 83

LC/MS analysis 84
HPLC assays 84
Protein crystallization and structure determination 85
Results and Discussion 93
Bioinformatic analysis of MT_L-GNAT initiation modules 93
AprA AR-MT_L-GNAT structure 95
Structural homology of MT_L 99
AprA MT_L catalyzes iron-dependent methyltransfer 101
Metal binding triggers tunnel formation 108
Substrate binding and catalysis 110
Ppant modeling 112
Relation of AprA MT_L to other PKS MTs 113
Functional annotation of GNAT initiation modules 113
Evolution of GNAT initiation modules 114
Chapter 5 Biosynthesis of *t*-Butyl in Apratoxin A: Functional Analysis and Architecture of a PKS Loading Module 116
Summary 116
Introduction 116
Experimental Procedures 119
Culturing, extraction, and purification of apratoxin A 119
NMR analysis and calculation of ^{13}C incorporation levels 121
Construct design 121
Protein expression and purification 126
Dimethylmalonyl-thiophenol synthesis 127
Production of acyl-ACPs 128
AprA $MT2_L$ enzyme assays 129
FabD enzyme assays 130
LC-MS analysis 130
HPLC analysis 130
Protein crystallization and structure determination 131
Negative-stain electron microscopy 133
Results and Discussion 133
Origin of the t-butyl group: SAM-derived methyl groups incorporated into apratoxin A 133

AprA MT2$_L$ is a dual function decarboxylase and methyltransferase 134
Crystal structure of AprA ΨGNAT-MT2$_L$ 136
Active site architecture 140
AprA MT2$_L$ is active on multiple substrates 140
Reaction mechanism for AprA MT2$_L$ 142
FabD provides the initial acyl transfer step 145
Modeling and EM visualization of the AprA module 148
Chapter 6 GNAT-like Domains for Initiation of Polyketide Biosynthesis are Primarily Acyl-ACP Decarboxylases 152
Summary 152
Introduction 152
Experimental Procedures 156
Construct design 156
Protein expression and purification 156
Production of acyl-ACPs 159
Decarboxylation enzyme assays 160
Acyl transfer enzyme assays 160
LC-MS analysis 161
Protein crystallization and structure determination 161
Results and Discussion 167
GNAT-like enzymes encode strict substrate specificity 167
GphF GNAT structure 168
GNAT active site and catalysis 170
GphF GNAT does not catalyze acyl transfer 174
PKS GNAT-like domains are acyl-ACP decarboxylases 180
Chapter 7 Conclusions and Future Directions 184
Overview 184
Structure and Activity of PKS MT Domains 184
α-methylation by *C*-MT domains 184
β-hydroxy methylation by *O*-MT domains 185
Malonyl methylation by MT$_L$ domains 186
t-Butyl production by MT2$_L$ 187
Functional Annotation of GNAT Containing Initiation Modules 187

Discovery of the ΨGNAT domain 187
Decarboxylation activity..... 188
Acyl transfer activity 189
Architecture of a PKS Initiation Module..... 190
Future Directions..... 190
Identification of novel natural products gene clusters..... 190
Engineering new polyketides and the development of biocatalysts..... 190
Understanding interdomain interactions and PKS module architecture..... 192
Evolutionary implications..... 193
Bibliography..... 195

List of Tables

Table 2.1 Primers for cloning *C*-MT constructs 21

Table 2.2 *C*-MT crystallographic information 23

Table 2.3 Scaling statistics for SeMet substituted CurJ *C*-MT (PDB ID 5THY) 24

Table 2.4 Scaling statistics for native CurJ *C*-MT (PDB ID 5THZ) 24

Table 2.5 Solubility of CurJ protein fragments 28

Table 2.6. Expected and observed mases of ACP species and Ppant ejection fragments 31

Table 2.7 Thermal stability of CurJ *C*-MT variants 42

Table 3.1 Primers for cloning *O*-MT constructs 47

Table 3.2 *O*-MT crystallographic information 53

Table 3.3 Scaling statistics for SeMet StiD 976-1266 (PDB ID 6ECU) 54

Table 3.4 Scaling statistics for StiD 976-1266 (PDB ID 6ECV) 54

Table 3.5 Scaling statistics for StiD 956-1266 (PDB ID 6ECW) 55

Table 3.6 Scaling statistics for StiE 961-1257 (PDB ID 6ECT) 55

Table 3.7 Scaling Statistics for StiE 942-1257 (PDB ID 6ECX) 56

Table 3.8 Protein stability of StiD fragments containing the *O*-MT 62

Table 4.1 Primers for subcloning MT_L-ΨGNAT 80

Table 4.2 AprA MT_L-ΨGNAT crystallographic information 87

Table 4.3 Scaling statistics for AprA MT_L-ΨGNAT SeMet metal free 88

Table 4.4 Scaling statistics for AprA MT_L-ΨGNAT metal free (PDB ID 6B39) 88

Table 4.5 Scaling statistics for AprA MT_L-ΨGNAT Mn bound (PDB ID 6B3A) 89

Table 4.6 Scaling statistics for AprA MT_L-ΨGNAT Mn and malonate bound (PDB ID 6B3B) . 89

Table 5.1 Enrichment of apratoxin A by culture of *M. bouillonii* PNG 5-198 supplemented with [methyl-^{13}C]methionine 122

Table 5.2 Enrichment of ^{13}C in apratoxin A by culture of *M. bouillonii* PNG 5-198 in media supplemented with [1-^{13}C]propionate 123

Table 5.3 ^{13}C-NMR shifts of native-abundance apratoxin A 124

Table 5.4 Primers for cloning $MT2_L$ and FabD constructs 125

Table 5.5 AprA $MT2_L$ crystallographic information 132

Table 6.1 Primers for cloning Gph and Cur GNAT constructs 157

Table 6.2 GphF GNAT crystallographic information 163

Table 6.3 Scaling statistics for GphF GNAT 164

Table 6.4 Scaling statistics for GphF GNAT in complex with isobutyryl-CoA 164

.

List of Figures

Figure 1.1 PKS, NRPS, and hybrid PKS/NRPS natural products 3

Figure 1.2 Biosynthesis of curacin A..... 4

Figure 1.3 PKS initiation mechanisms..... 5

Figure 1.4 PKS catalytic cycle..... 7

Figure 1.5 Comparison of mFAS and PKS architecture..... 9

Figure 1.6 Class I MT core and conserved motifs 11

Figure 1.7 The GNAT-like family of initiation modules..... 14

Figure 2.1 Introduction of α-methyl by CurJ *C*-MT. 19

Figure 2.2 Ramachandran plots for CurJ *C*-MT SeMet..... 25

Figure 2.3 Ramachandran plots for native CurJ *C*-MT 26

Figure 2.4 Sequence alignment of PKS *C*-MTs with CurJ *C*-MT secondary structure annotation 29

Figure 2.5 Dendogram of PKS *C*-MT domains. 30

Figure 2.6 LC-MS analysis of *C*-MT activity. 32

Figure 2.7 Mass spectra of CurJ *C*-MT reactions on NAC substrates..... 33

Figure 2.8 Representative data from acetoacetyl-ACP Ppant ejection assay 34

Figure 2.9 CurJ *C*-MT structure..... 36

Figure 2.10 Comparison of CurJ *C*-MT and mFAS ΨMT 37

Figure 2.11 Architecture of PKS and mFAS modifying regions. 38

Figure 2.12 Omit density for SAH 40
Figure 2.13 CurJ *C*-MT active site. 41
Figure 2.14 Movement surrounding the CurJ *C*-MT active site. 41
Figure 3.1 Reactions carried out by PKS *O*-MTs 46
Figure 3.2 Ramachandran plots for SeMet labeled StiD *O*-MT 976-1266 57
Figure 3.3 Ramachandran plots for native StiD *O*-MT 976-1266 58
Figure 3.4 Ramachandran plots for StiD *O*-MT 956-1266 59
Figure 3.5 Ramachandran plots for StiE *O*-MT 961-1257 60
Figure 3.6 Ramachandran plots for StiE *O*-MT 942-1257 61
Figure 3.7 Sequence alignment of PKS *O*-MTs 63
Figure 3.8 Oligomeric state of *O*-MTs 64
Figure 3.9 Activity of StiD *O*-MT on ACP linked substrates 65
Figure 3.10 Activity of StiE and CurL *O*-MTs on ACP linked substrates 66
Figure 3.11 StiD and StiE *O*-MT structures. 67
Figure 3.12 StiD and StiE *O*-MT dimer interfaces 68
Figure 3.13 StiD lid dynamics 69
Figure 3.14 PKS *O*-MT homologs 70
Figure 3.15 StiE active site density 71
Figure 3.16 Relative methylation activities of CurL and StiE *O*-MT variants 73
Figure 3.17 Conservation of key active site residues in *O*-MT homologs 74
Figure 4.1 Domain architecture of GNAT containing initiation modules 78
Figure 4.2 Ramachandran plots for AprA MT_L-ΨGNAT metal free 90
Figure 4.3 Ramachandran plots for Mn bound AprA MT_L-ΨGNAT 91

Figure 4.4 Ramachandran plots for Mn and malonate bound AprA MT_L-ΨGNAT 92

Figure 4.5 AR-MT1 sequence alignment with AprA MT_L secondary structure annotation 94

Figure 4.6 Structures of AprA AR-MT_L-ΨGNAT colored by structural region 95

Figure 4.7 AprA MT_L active sites 96

Figure 4.8 Omit density AprA MT_L active sites 97

Figure 4.9 Superposition of AprA ΨGNAT and CurA GNAT 98

Figure 4.10 GNAT sequence alignment with CurA GNAT secondary structure annotation 99

Figure 4.11 MT_L and *C*-MT homology 100

Figure 4.12 AprA MT_L-ΨGNAT activity. 102

Figure 4.13 Representative electrospray-injection (ESI) mass spectra of AprA ACP from MT_L reaction mixes 103

Figure 4.14 Time course data for AprA MT_L-ΨGNAT methylation reactions 104

Figure 4.15 AprA MT_L-ΨGNAT reactions with Mal-ACP, Mal-CoA or MeMal-CoA analyzed by HPLC 105

Figure 4.16 GphF MT_L-GNAT reaction with Mal-ACP 106

Figure 4.17 Mass spectra of methylation and acetyl transfer reactions with AprA MT_L-ΨGNAT, AprA ΨGNAT, and CurJ *C*-MT 107

Figure 4.18 Key amino acids for substrate binding 108

Figure 4.19 Relative methylation activities of wild type AprA MT_L-ΨGNAT, serendipitous crystallization substitution (S274I/Q528P), and active site variants 109

Figure 4.20 AprA MT_L mechanism 111

Figure 4.21 Malonyl-Ppant substrate modeling 112

Figure 5.1 Production of a *t*-butyl group by AprA 118

Figure 5.2: Comparison of ^{13}C-NMR enrichment of apratoxin A with methionine and propionate 120

Figure 5.3 Catalytic activity of $MT2_L$ 135

Figure 5.4 Representative mass spectra for the production of Me_2Mal-ACP 136

Figure 5.5 AprA ΨGNAT- $MT2_L$ structure and active site 137

Figure 5.6 AprA $MT2_L$ active site omit density and comparison between AprA $MT2_L$ and CurJ *C*-MT 138

Figure 5.7 AprA $MT2_L$ sequence alignment and secondary structure with BryX $MT2_L$ and PKS *C*-MTs. 139

Figure 5.8 AprA $MT2_L$ cofactor dependence 141

Figure 5.9 Probing AprA $MT2_L$ activity via site-directed mutagenesis 143

Figure 5.10 Representative mass spectra for coupled decarboxylation and methylation reactions 144

Figure 5.11 Acyltransfer activity and GNAT initiation-module-associated ACPs 146

Figure 5.12 FabD malonyl acyltransfer activity 148

Figure 5.13 Architecture of the AprA module 149

Figure 5.14 EM class averages, micrographs and AprA model 150

Figure 6.1 GNAT-like enzymes in gephyronic acid and curacin A biosynthesis. 154

Figure 6.2 Sequence alignment of GNAT-like domains from PKS pathways for natural products of known structure 155

Figure 6.3 Ramachandran plots for GphF GNAT 165

Figure 6.4 Ramachandran plots for GphF GNAT in complex with isobutyryl-CoA 166

Figure 6.5 GphF and CurA GNAT decarboxylation activity 168

Figure 6.6 Representative mass spectra of decarboxylation reactions 169

Figure 6.7 GphF GNAT structure 170

Figure 6.8 Probing decarboxylation via site directed mutagenesis. 171

Figure 6.9 Active site pockets of GNAT-like enzymes and modeled Me_2Mal-CoA complex... 173

Figure 6.10 GphF and CurA GNAT acyl transfer assays 176

Figure 6.11 Representative mass spectra of GphF GNAT and MT-GNAT acyl transfer reactions 177

Figure 6.12 Decarboxylation by extensively purified GphF and CurA GNAT used for acyl transfer assays 178

Figure 6.13 Representative mass spectra of CurA GNAT and *M. bouillonii* FabD acyl transfer reactions 179

Figure 6.14 Sequence alignment of the GNAT domain of EryM / SACE_1304 / Mcd homologs and Rv1347c. 181

Abbreviations

ACP, acyl carrier protein

ACP_L, initiation module acyl carrier protein

ANS, 1-anilinonaphthalene-8-sulfonic acid

Apr, apratoxin A biosynthetic pathway

APS, advanced photon source

AR, adaptor region

AT, acyltransferase

ATCC, American type culture collection

AT_L, initiation module acyltransferase

Bat, batumin biosynthetic pathway

BCDH, branched chain α-keto acid dehydrogenase

Bon, bongkrekic acid biosynthetic pathway

Bry, bryostatin biosynthetic pathway

C-MT, carbon methyltransferase

CoA, coenzyme A

Crp, cryptophycin biosynthetic pathway

Cur, curacin A biosynthetic pathway

Da, Dalton

DEBS, erythromycin biosynthetic pathway

DSC, differential scanning calorimetry

EM, electron microscopy

ER, enoylreductase

Ery, erythromycin biosynthetic pathway

ESI, electrospray ionization

FabD, bacterial fatty acid synthase malonyl acyltransferase

GC-MS, gas chromatography-mass spectrometry

GCN5, general control non-repressible 5

GNAT, GCN5-related *N*-acetyltransferase

Gph, gephyronic acid biosynthetic pathway

GSH, glutathioneGSSH, glutathione disulfide

HPLC, high performance liquid chromatography

IPTG, isopropyl β-D-1-thiogalactopyranoside

Jam, jamaicamide biosynthetic pathway

KR, ketoreductase

KR_c, ketoreductase catalytic domain

KR_s, ketoreductase structural domain

KS, ketosynthase

KS_L, initiation module decarboxylating ketosynthase

LC-MS, liquid chromatography, mass spectrometry

LIC, ligation independent cloning

Lov, lovastatin biosynthetic pathway

Mal-, malonyl

MAS, mycocerosic acid

MCD, malonyl-CoA decarboxylase

Me_2Mal-, dimethylmalonyl

MeMal-, methylmalonyl

mFAS, metazoan fatty acid synthase

Mpp, mannopeptimycin

MS, mass spectrometry

MT, methyltransferase

$MT2_L$, second initiation module methyltransferase

MT_L, initiation module methyltransferase

NAC, *N*-acetyl cysteamine

NMR, nuclear magnetic resonance spectroscopy

N-MT, nitrogen methyltransferase

NRPS, non-ribosomal peptide synthetase

Nsp, nosperin biosynthetic pathway

O-MT, oxygen methyltransferase

Onn, onnamide A biosynthetic pathway

PCR, polymerase chain reaction

Ped, pederin biosynthetic pathway

PKS, polyketide synthase

Plm, phoslactomycin biosynthetic pathway

Ppant, phosphopantetheine

QTOF, quadrapole time of flight

Rhi, rhizoxin biosynthetic pathway

RiPP, ribosomally synthesized and post-translationally modified peptides

RMSD, root-mean-square deviation

SAD, single-wavelength anomalous diffraction

SAH, *S*-adenosyl-L-homocysteine

SAM, *S*-adenosyl-L-methionine

SeMet, selenomethionine

Spn, spinosyn biosynthetic pathway

Sti, stigmatellin biosynthetic pathway

SVP, *Streptomyces verticillus* phosphopantethinyl transferase

TaI, myxovirescin A biosynthetic pathway

TCEP, Tris(2-carboxyethyl) phosphine

TE, thioesterase

TEV, tobacco etch virus

ΨGNAT, vestigial pseudo GNAT domain

ΨMT, vestigial pseudo methyltransferase

Abstract

Polyketide natural products are a chemically diverse class of small molecules possessing a variety of therapeutic applications. Modular type I polyketide synthases (PKS) use a series of multienzyme modules in the assembly-line synthesis of polyketides from coenzyme A (CoA) building blocks. Biosynthetic intermediates are covalently linked to modules through an acyl carrier protein (ACP). The chemical diversity of polyketides is achieved through the variety of the catalytic domains within each module. In order to harness the biocatalytic power of PKS for the production of novel molecules, it is essential to understand the structural and mechanistic details of each biosynthetic tool. Through the use of x-ray crystallography and biochemical assays, this thesis investigates the role of methyltransferases (MTs), the least studied modification domain, in polyketide biosynthesis.

Four distinct methyltransferases are found in PKS pathways: carbon (*C*-) and oxygen (*O*-) MTs occur in polyketide extension modules; two other MT types (MT_L and $MT2_L$) are exclusive to "loading" modules, which initiate PKS biosynthesis. Biochemical studies divulged the substrate for the *C*-MT, the first *C*-MT crystal structure revealed its common ancestry with the vestigial pseudo-MT of metazoan fatty acid synthase (mFAS). The PKS *C*-MTs and *O*-MTs were found to arise from different branches of the MT superfamily. Identification of essential catalytic residues for *C*-MTs and *O*-MTs provides insight into the methylation mechanism.

A new biosynthetic route to a *t*-butyl group was a major discovery of this thesis. MT_L and $MT2_L$ are associated with initiation modules that also contain GNAT-like acyltransferase/decarboxylase domains. Branched chain propionyl and isobutyryl starter units are generated by modules containing MT_L, whereas MT_L and $MT2_L$ together synthesize a *t*-butyl group. The AprA MT_L from the apratoxin A biosynthetic pathway was discovered to be a rare iron-dependent MT, which converts malonyl-ACP to dimethylmalonyl-ACP through a methylmalonyl-ACP intermediate. In contrast, the AprA $MT2_L$, a homolog of PKS *C*-MT domains, is a bifunctional enzyme that catalyzes coupled decarboxylation and methylation reactions to directly convert dimethylmalonyl-ACP to the *t*-butyl-containing pivaloyl-ACP. The

AprA module was further visualized by negative-stain electron microscopy, revealing a dynamic module that may exist in different states for the MT_L and $MT2_L$ catalytic steps.

Analysis of MT_L prompted further investigation of the GNAT-like domains in propionyl- and isobutyryl-ACP producing modules. Characterization of the GphF GNAT from the gephyronic acid biosynthetic pathway, which produces isobutyryl-ACP, demonstrated that the GNAT domain acts as a gatekeeper, selectively decarboxylating the MT_L methylation product (dimethylmalonyl-ACP) for further processing by the enzymatic assembly line. Surprisingly, the expected acyl transfer activity was not detected, prompting the reclassification of PKS GNAT-like domains as acyl-ACP decarboxylases. The bacterial FAS malonyl-acyltransferase was investigated as a candidate for the acyltransfer reaction. The FAS malonyl-acyltransferase supports the initial acyl transfer step to prime the loading module ACP, potentially linking primary and secondary metabolism in the producing organism. Characterization of PKS MTs and the acyl-ACP decarboxylases advances our understanding of the biosynthesis of many valuable natural products and provides initial tools for the development of biocatalysts capable of synthetically challenging stereo- and regiospecific methylation reactions.

Chapter 1 Introduction

Natural Products

Secondary metabolites are molecules produced by living organisms that are not essential for survival, but often confer a selective advantage towards survival. For example, secondary metabolites provide protection by inhibiting the growth of competing organisms[6], acting as feeding deterrents[7], or scavenging precious resources from the environment required for cell growth[8]. Overall, the secondary metabolome acts as a chemical defense mechanism for organisms lacking a sophisticated immune system (e.g. bacteria, fungi, plants)[9]. For centuries, secondary metabolites have been repurposed as "natural products", as they often possess medicinally valuable bioactivities[10]. Currently, a large proportion of clinical pharmaceuticals are derived from or inspired by natural products[11]. However, the chemical complexity of natural products renders them challenging for structural elucidation and chemical synthesis, making the lack of access to compounds at sufficient quantities from both synthetic and biological sources an impediment for the development of pharmaceuticals. Over the past two decades natural products have fallen out of style in drug discovery efforts in favor of high throughput screening of synthetic small molecule libraries[12]. As this strategy has not yielded new drugs to combat the bacterial antibiotic resistance crisis, interest is once again shifting toward natural products as a source of new pharmaceuticals[13].

Biosynthetic Gene Clusters

Advances in genomics and the discovery that many secondary metabolites are produced by genes located in continuous stretches in bacterial genomes[14] has rapidly linked natural products with the enzymes responsible for their production. Identification of the biosynthetic gene cluster for a target molecule can facilitate the production of a natural product in a genetically tractable and readily culturable heterologous host. Genome sequencing has led to the bioinformatic discovery of new "cryptic" biosynthetic pathways for previously undetected molecules, yielding new natural products for drug discovery[15]. Additionally, gene cluster mining has provided critical

insight into the evolution of natural product biosynthetic pathways, which often occurs through horizontal gene transfer, mutation of individual genes, and gene duplication[16]. Finally, identification of the biosynthetic tools responsible for synthesizing some of the most chemically complex molecules in nature provides valuable biocatalysts for synthetically difficult transformations[17].

The products of polyketide synthase (PKS), non-ribosomal peptide synthetase (NRPS), and hybrid PKS/NRPS are among the most prevalent secondary metabolites. The two megasynthase types synthesize natural product with exquisite stereo and regioselectivity via a series of enzymes. Polyketides are produced through the successive condensation of acyl-CoA building blocks to produce chemically diverse molecules, such as the antibiotic erythromycin[18], electron transport inhibitor stigmatellin[19], protein kinase C modulator bryostatin[20], and cytostatic agent gephyronic acid (Figure 1.1)[21]. NRPS use different enzymatic machinery to condense amino acids, often forming cyclic peptides, with a prominent example being the clinically essential immunosuppressant cyclosporin (Figure 1.1)[22]. Secondary metabolites produced by hybrid PKS/NRPS pathways include the antimitotic curacin A[23], neurotoxin jamaicamide[24] and apratoxin A[25], an inhibitor of the protein secretory pathway (Figure 1.1)[26, 27].

Type I Polyketide Synthases

All PKSs use acyl-CoA building blocks, but are classified by their pathway organization, sequences, and catalytic mechanisms. In type I PKS[28], polyketides are synthesized by a multifunctional enzyme, with the substrates attached to a phosphopantetheine (Ppant) prosthetic group of an acyl carrier protein (ACP) domain via a thioester linkage. The ACP shuttles the substrates between the various enzymatic domains. Iterative type I PKS use a single polypeptide to synthetize the polyketide through multiple additions of an acyl building block, whereas modular type I PKS are composed of a multimodule assembly line with each module responsible for the successive addition of an acyl-CoA building block (Figure 1.2). Modular type I PKS can be further divided into *cis*-AT and *trans*-AT PKS, in which the acyltransferase domain (AT) responsible for selecting and transferring the acyl building block from CoA to the ACP is found either within each module (*cis*-AT) or as a single stand-alone enzyme (*trans*-AT) for all modules in the pathway. Catalytic steps of type II PKS are carried out by multiple discrete enzymes, with the substrates and products tethered to an ACP[29]. A single enzyme carries out multiple catalytic steps in type III PKS, which are mainly carrier protein independent[30].

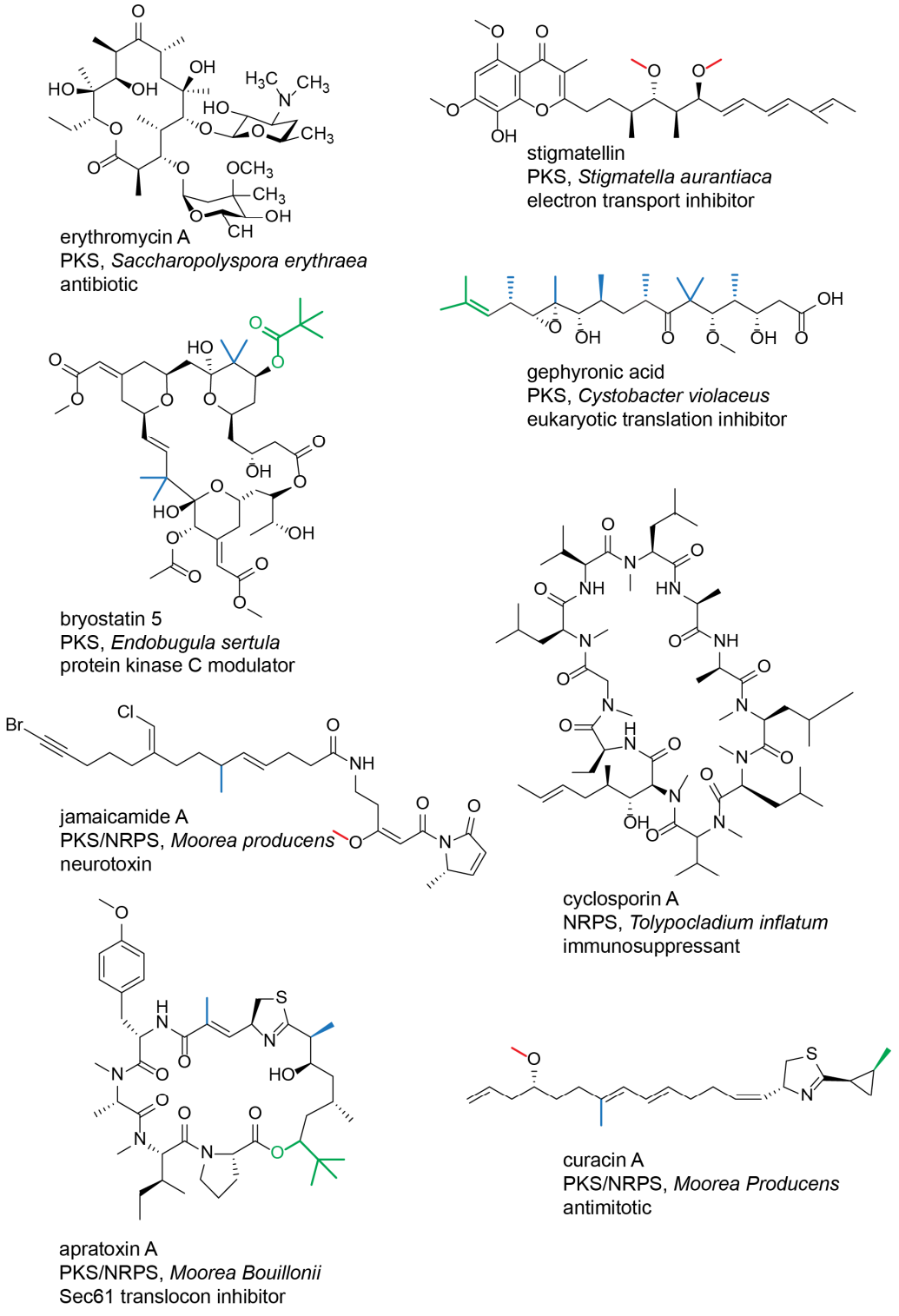


Figure 1.1 PKS, NRPS, and hybrid PKS/NRPS natural products
Chemical structures of bioactive natural products. Biosynthetic system, producing organism, and bioactivity are listed below each molecule. Colored bonds represent functionalities installed by enzymes studied in this thesis: blue, PKS *C*-methyltransferases (*C*-MTs); red, PKS *O*-MTs; green, products of initiation modules containing GNAT-like enzymes.

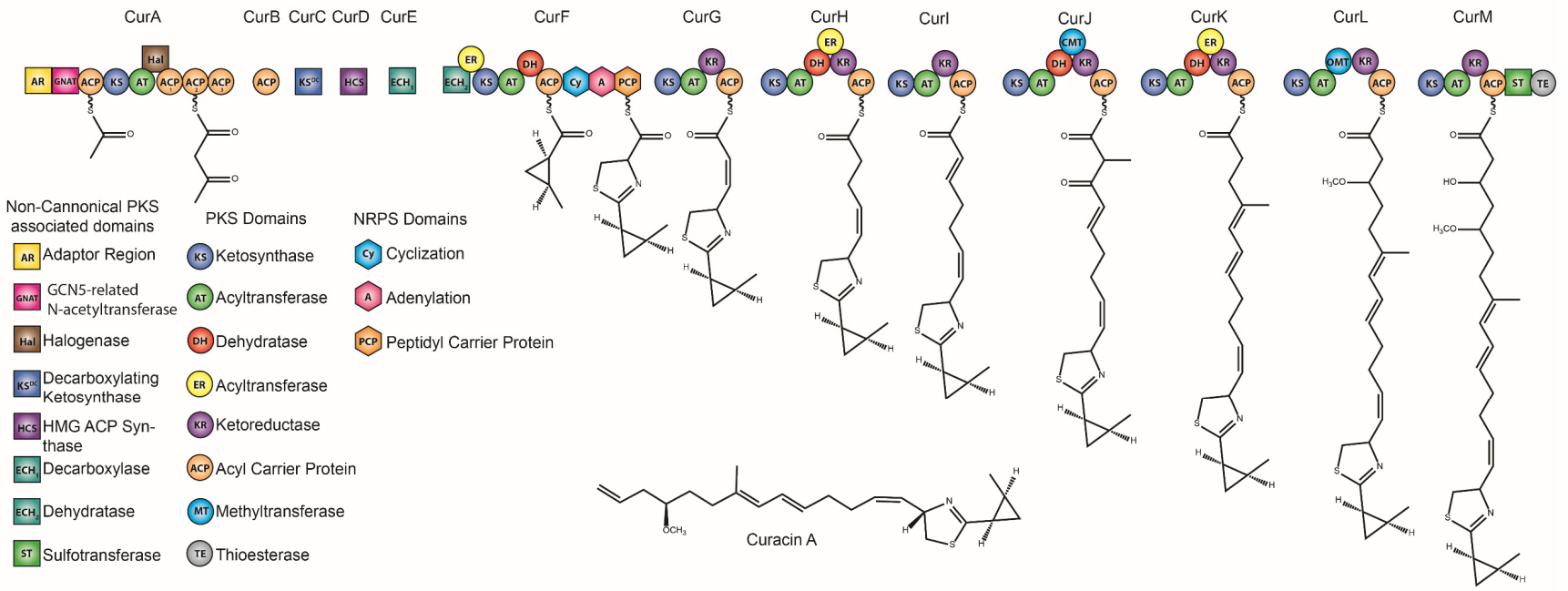


Figure 1.2 Biosynthesis of curacin A

Curacin A is synthesized by a hybrid multimodular PKS/NRPS assembly line. Canonical PKS domains are shown in circles, NRPS modules are shown in hexagons. Non-canonical enzymes found in some PKS pathways are shown in squares. The CurA initiation module contains a GNAT-like domain which decarboxylates malonyl-CoA to acetyl-CoA and performs an *S*-acetyltransfer reaction yielding acetyl-ACP[31]. Acetyl-ACP is converted to a methyl cyclopropyl moiety[32] through a KS-AT extension reaction (CurA), the action of β-branching enzymes (ACP, CurB; KS_{DC}, CurC; HCS, CurD)[33], a halogenase (Hal, CurA)[34], two enoyl-CoA hydratases (ECH_1, CurE; ECH_2, CurF)[35], and an enoyl reductase (ER, CurF) that carries out the cyclopropanation reaction[36]. An NRPS module (Cy-A-PCP, CurF) then appends a thiazole, derived from cysteine, to the cyclopropyl[37]. The product of CurF is shuttled to a canonical modular type I PKS module CurG, where the first of seven PKS extension reactions occurs (CurG-CurM). At the end of the assembly line, a terminal alkene is formed through the action of a sulfotransferase (ST) and thioesterase (TE)[38-40].

PKS pathway initiation

Synthesis by type I PKS is initiated by the transfer of an acyl group from CoA to the first ACP in the pathway, which is located in the "loading" or "initiation" module. The initial acyltransfer step can occur through several mechanisms. Most frequently, the loading module contains a loading acyltransferase (AT_L), which selects a specific acyl-CoA substrate, and transfers the acyl group to the loading ACP (ACP_L) (Figure 1.3). Carboxylated substrates, such as malonyl and methylmalonyl can be processed by a specialized decarboxylating ketosynthase domain (KS_L) encoded in the initiation module (Figure 1.3)[41].

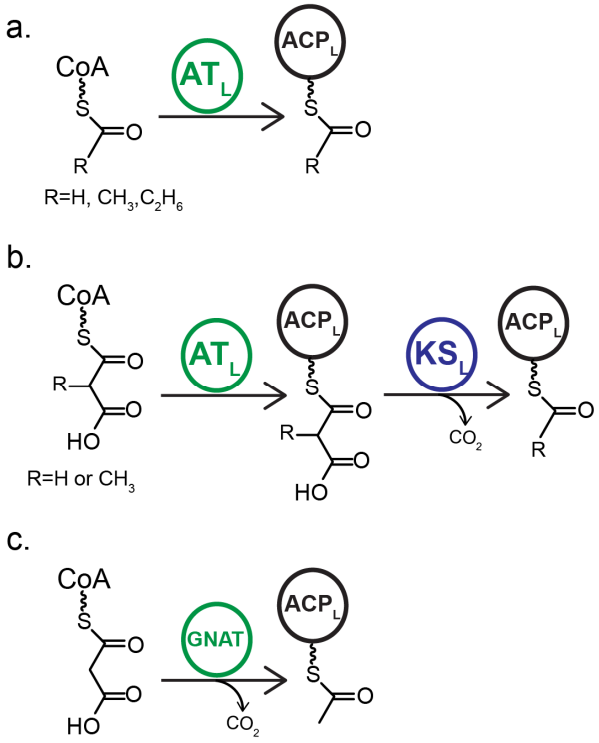


Figure 1.3 PKS initiation mechanisms
Most PKS initiation modules contain a loading acyltransferase (AT_L) which **a.** selects non-carboxylated CoAs (acetyl, propionyl, isobutyryl) and directly loads the acyl group onto the ACP or **b.** selects a carboxylated CoA (malonyl, methylmalonyl), which is decarboxylated by a specialized ketosynthase (KS_L) that carries out decarboxylation only[41]. **c.** The GNAT-like family of initiation modules contains a GNAT domain that carries out the functions of both AT_L and KS_L in a single domain[31].

A second mechanism of pathway initiation uses a loading module containing a GCN5-related acyltransferase-like (GNAT-like) domain (Figure 1.3). Members of the GNAT superfamily typically carry out an *N*-acetyltransfer reaction between acetyl-CoA and a variety of small molecule and protein amine substrates[42, 43]. The PKS GNAT-like domain from the pederin biosynthetic pathway was the first identified member of the GNAT superfamily in PKS based on sequence analysis[44]. The existence of a phosphate-binding P-loop sequence, also observed in the acetyl-CoA binding GNATs, suggested that the pederin GNAT-like domain shares the canonical acetyl-CoA binding property of the GNAT superfamily and provides an acetyl-ACP starter unit to the pederin pathway through an *S*-acetyltransfer reaction. The *S*-acetyltransfer activity was experimentally validated in the CurA GNAT-like enzyme from the curacin A pathway. However, a surprising malonyl-CoA decarboxylation activity was also discovered for CurA GNAT[31]. Therefore, PKS GNAT-like domains are proposed to decarboxylate malonyl-CoA and then transfer the resulting acetyl group from CoA to ACP. Acetyl-ACP is hypothesized to be further processed by methyltransferases (MTs) encoded in the initiation module[45].

Polyketide chain extension and modification

The acyl-ACP_L product of the initiation module (**3**, Figure 1.4) is passed to the ketosynthase (KS) of the first extension module of the pathway, where it is transferred from the ACP_L Ppant to the KS active site cysteine via a transthioesterification forming an acyl enzyme intermediate. Malonyl- or methylmalonyl-ACP (**2**, Figure 1.4) is provided by the extension module AT domain, and the KS then catalyzes a decarboxylative Claisen condensation between the carboxylated building block and KS acyl enzyme intermediate extending the polyketide by two carbons. The resulting polyketide intermediate (**5**, Figure 1.4) contains a keto group at the β-carbon position. The formation of a new carbon-carbon bond by the KS is the minimal reaction carried out by a PKS module containing KS-AT and ACP domains. Additional processing of the β-keto intermediate can be catalyzed by a variety of optional modification domains within the PKS module. The β-keto intermediate (**5**, Figure 1.4) can be stereo-selectively reduced by the NADPH-dependent ketoreductase domain producing a β-hydroxyl intermediate (**7**, Figure 1.4), which can be further reduced by a dehydratase (DH) to produce a *cis*-or *trans*-α,β-double bond (**10**, Figure 1.4). Action of an NADPH-dependent enoylredutase (ER) domain produces a fully reduced polyketide intermediate (**11**, Figure 1.4). Non-reductive modification domains include MTs. Once the nascent polyketide has been processed by each domain in the module, the ACP travels to the

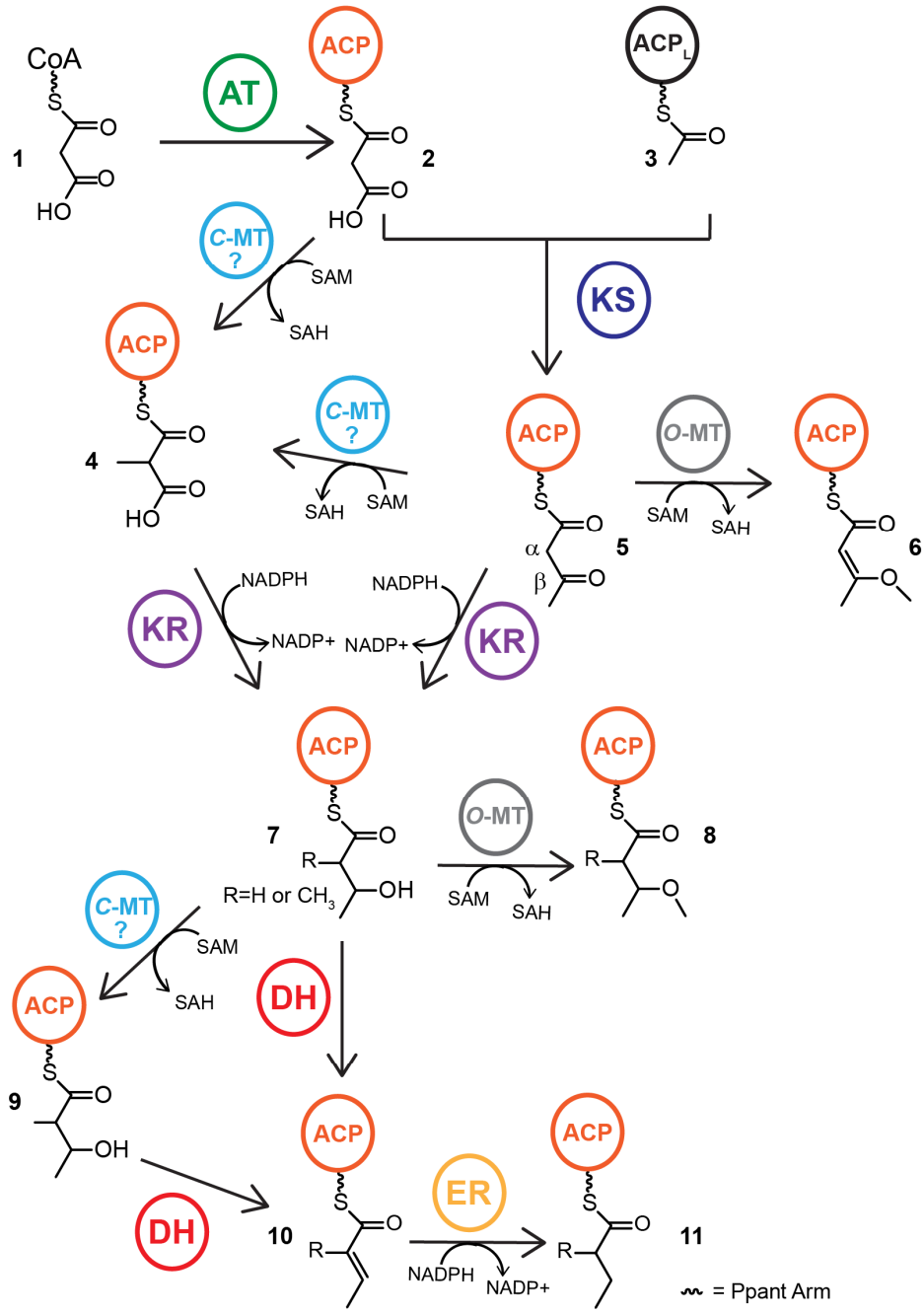


Figure 1.4 PKS catalytic cycle

The AT domain (green) selects the building block malonyl-CoA (**1**) and delivers it to the ACP (orange) Ppant arm (**2**). The upstream initiation module ACP (ACP_L) (black) delivers the starter unit (**3**) to the KS (blue), where it is combined with **2** to extend the polyketide by two carbons (**5**). **5** is reduced by the KR (purple) domain to form the β-hydroxy intermediate (**7**). In modules containing a *C*-MT (cyan), it is unclear whether α-methylation takes place on **2**, **5**, or **7**. **7** can be dehydrated by the DH (red, **10**) and reduced by the ER (yellow, **11**) domains. Modules containing *O*-MTs (gray) methylate the β-keto of **5** or the β-hydroxyl of **7**. Depending on the series of modification domains encoded in the module, **5**, **6**, **7**, **8**, **9**, **10**, and **11** can be passed onto the next extension module in the pathway.

next extension module for additional rounds of polyketide chain extension and modification. Modules act in a prescribed order, interacting through small docking domains found at the N- and C-termini of the multienzyme polypeptides[46-48]. The domain content of each module stipulates the chemistry carried out after each chain extension reaction (Figure 1.2). Variation in module's domain content yields chemically diverse polyketides from common enzymatic machinery. At the end of the assembly line, the polyketide is removed from the ACP by a thioesterase (TE) domain that catalyzes hydrolysis or cyclization (Figure 1.2). The polyketide core scaffold is often glycosylated, methylated, or modified by cytochrome P450 enzymes to produce the final bioactive natural product[49].

PKS Module Architecture and Relationship to Mammalian Fatty Acid Synthase

PKS modules are obligate dimers. Dimerization of each module is mediated by a dimeric KS domain (Figure 1.5)[48, 50, 51] along with small dimerization elements, composed of helices found after the AT domain[52] or ACP[46, 53] domain. The oligomeric state of modules can be additionally stabilized by optional dimeric DH[54, 55] and TE[56, 57] domains. Excised KR[2, 4, 58] and ER[36] domains are monomeric. The monomeric KR domain is composed of two subdomains, a N-terminal structural domains (KR_s) and C-terminal catalytic domain (KR_c), that appear to be a result of a gene duplication event. When present, the ER domain is inserted in between the KR_s and KR_c[5].

Polyketide biosynthesis is analogous to the biosynthesis of fatty acids in primary metabolism. Type I PKS are most similar to metazoan fatty acid synthase (mFAS), which is also an iterative multienzyme megasynthase[59]. mFAS and PKS contain the same enzymatic domains that perform identical chemical transformations to produce their respective polyketide and fatty acid products. Additionally, the shared enzymatic domains are found in the same order on mFAS and PKS polypeptides. Like PKS, the mFAS KR is composed of a KR_s and KR_c. The crystal structure of an mFAS (Figure 1.5) revealed an additional domain embedded in KR_s that has structural homology to a MT but lacks the characteristic sequence motifs for binding of the methyl donor *S*-adenosylmethionine (SAM). Thus, this domain was classified as an inactive vestigial "pseudo" MT (ΨMT) [1]. Interestingly, some PKS modules contain active MT domains adjacent to the KR_s.

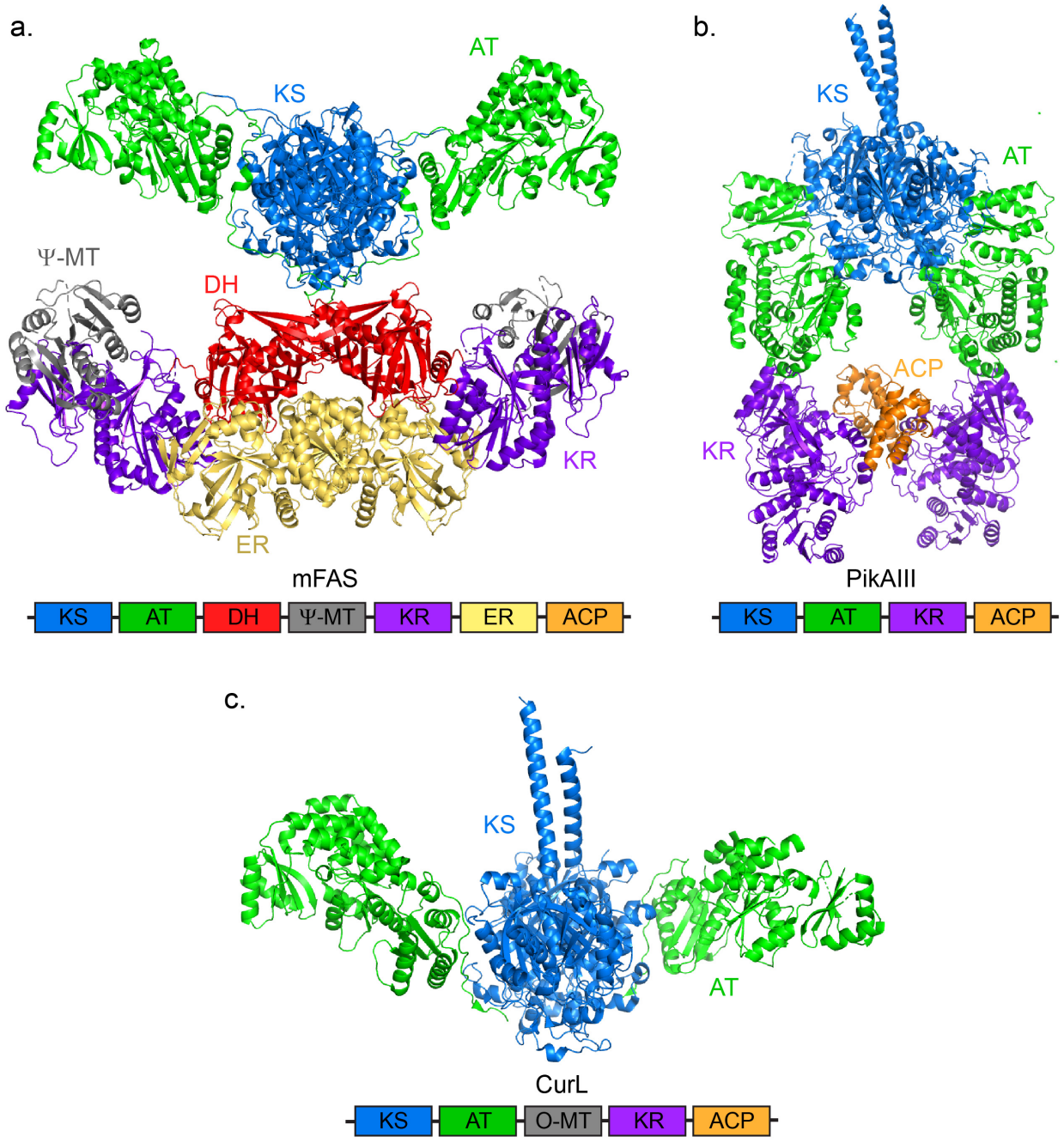


Figure 1.5 Comparison of mFAS and PKS architecture
Structures of **a.** mammalian fatty acid synthase (mFAS) (PDB ID 2VZ8)[1], **b.** PikAIII[53], **c.** CurL KS-AT fragment (PDB ID 4MZ0)[48]. Domains are colored as follows: ketosynthase (KS), blue; acyltransferase (AT), green; dehydratase (DH), red; pseudo-methyltransferase (ΨMT) or oxygen-methyltransferase (*O*-MT), gray; ketoreductase (KR), purple; enoylreductase (ER), yellow; acyl carrier protein (ACP), orange. Domains within the full-length module are depicted below the structures.

While representative crystal structures have been solved for nearly all PKS domains[28], information regarding domain-domain interactions, which may hint to the overall architecture of PKS modules, is limited to several KS-AT crystal structures[3, 48, 50, 51], a KR-ER di-domain structure[5], and a recent structure of the DH-KR-ER tri-domain from the mycocerosic acid synthase (MAS)-like iterative PKS[3]. For years, the crystal structure of mFAS was the model for PKS architecture, primarily based on similarities between the mFAS and PKS KS-AT didomain crystal structures (Figure 1.5). However, differences in interdomain linker lengths, the oligomeric state of excised domains, and the inherent ability of PKS modules to gain or lose modification domains, hints that PKS and mFAS could have architectural differences. Structural characterization of the first intact KS-AT-KR-ACP PKS module PikAIII in defined chemical states revealed a drastically different architecture compared to mFAS (Figure 1.5)[53, 60]. Instead of the splayed shape of mFAS, PikAIII adopts an arched architecture, creating a central reaction chamber that sequesters the ACP-bound intermediates during the PKS catalytic cycle. However, the moderate resolution (7-8 Å) of the EM reconstructions limits interpretation of interdomain interactions at an atomic level. It remains unclear if PKS modules containing the same domains as mFAS or other combinations of modification domains will more closely resemble mFAS or PikAIII.

Methyltransferases in Natural Product Biosynthesis

The 2000-plus MT structures deposited in the structural database represent 15 different superfamilies, 11 of which require the methyl donor SAM. SAM-dependent MTs are found throughout all kingdoms of life and catalyze over 90% of the 300-plus methylation reactions in biology[61]. Class I MTs are the most predominant MT superfamily, characterized by a seven-β-stranded core containing six conserved motifs (Figure 1.6)[62]. They typically methylate *O*-, *N*-, *C*, and *S*-nucleophiles through an acid-base or metal-dependent mechanism. Although class I MTs share a common core, the SAM cofactor or demethylated product *S*-adenosylhomocysteine (SAH) is observed in a variety of conformations, which may correlate to different branches of the class I MT family[63]. Branches of the class I MT family are distinguished by additional structural elements appended to the common core, often facilitating substrate binding or protein oligomerization. However, branch members share low sequence identity (<20%) to each other as well as members of other branches. The wide variety of methylation reactions and substrates accommodated by the

common SAM binding core makes class I MTs prime candidates for the development of biocatalysts[64].

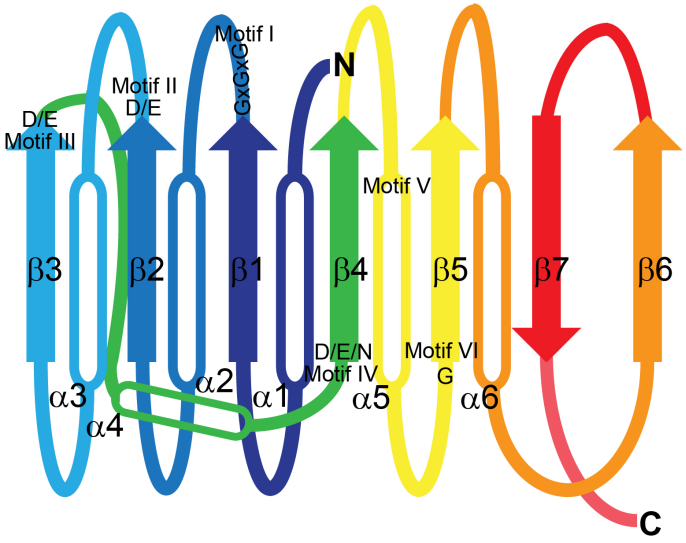


Figure 1.6 Class I MT core and conserved motifs
All class I MTs share a seven-β-stranded core. Six conserved motifs (labeled above in roman numerals) occur throughout the core[62]. SAM generally binds close to Motifs I, II and III. Motif I is essential for SAM binding. Motifs II and III contain acidic residues at the C-termini of their respective β-strands. Motif IV contains a D/E/N residue. Motif V often contains a hydrophobic residue that interacts with the SAM adenine ring. Motif VI contains an invariant glycine in atight turn preceding β5. Branches of the class I MT family are distinguished by insertions at the N- and C-termini or within core loops.

Natural product biosynthetic pathways contain a plethora of small molecule MTs, most of which belong to the class I fold[65]. Oxygen methyltransferases (*O*-MTs) are often encoded in natural product pathways and frequently methylate sugars[66-68] or phenols[69, 70]. Some *O*-MTs require metal cofactors, typically Mg^{2+} or Ca^{2+}, for positioning and activation of the hydroxyl group substrate[66, 67]. Surprisingly, higher organisms, such as yeast and humans, contain only two *O*-MTs that methylate small molecule substrates[65]. Similarly, small molecule carbon methyltransferases (*C*-MTs) are much more common to bacteria and plants than animals[65]. Examples include *C*-MTs that methylate coumarin scaffolds[71], terpenoids[72], tetrapyrroles[73], and amino acid derivatives[74]. Like *O*-MTs, some *C*-MTs require metals to position the substrate[72] or facilitate deprotonation of the target carbon[74]. Finally, natural product *N*-methyltransferases (N-MTs) frequently methylates N-terminal[75, 76] or backbone amides[77-79] of amino acids in NRPS and

ribosomally synthesized and posttranslationally modified peptide (RiPP) natural products. *N*-MTs are also commonly found encoded as stand-alone enzymes in PKS gene clusters, where they methylate amines on sugars that decorate the polyketide macrolide core [80, 81].

PKS C-MT domains

PKS modules can encode a ~390 amino acid *C*-MT domain (Figure 1.2), which is a member of the class I MT superfamily and catalyzes the addition of a methyl or gem-dimethyl moiety at the α-position of the growing polyketide (Figure 1.1, Figure 1.4). Modification by the *C*-MT and incorporation of a methylmalonyl building block, result in the identical α-methyl functionality. Initially identified and studied in the fungal lovastatin PKS pathway[82] and yersiniabactin hybrid PKS-NRPS pathway[83], *C*-MTs have been annotated in PKS modules from fungal iterative PKS[84-88] as well as bacterial modular *cis*-AT[21, 23-25, 89-92] and numerous *trans*-AT PKS[20, 93-98]. In iterative and *cis*-AT PKS the *C*-MT domain is embedded in the KR domains, following a conserved β-strand linker, which marks the start of the KR_S domain, but before the KR_S core, whereas in *trans*-AT PKS the *C*-MT is found after additional modification domains, preceding the ACP. Although *C*-MTs are well annotated in pathway databases, their structural and biochemical characterization has lagged behind the reductive modification domains. Domain boundaries for the *C*-MTs are unclear, especially at the *C*-terminus where sequence conservation is low, forcing most biochemical interrogations of PKS *C*-MT activity to be carried out in intact modules[84, 99-101].

The timing of α-methyl installation by the *C*-MT within the PKS catalytic cycle is ambiguous (Figure 1.4), as methylation could occur on the malonyl-ACP building block (**2**, Figure 1.4) prior to KS condensation reaction, on the β-keto product of the condensation reaction (**5**, Figure 1.4), or even on a product of a reductive modification domain. Studies on fungal PKS suggest that the β-keto intermediate (**5**, Figure 1.4) is the substrate for α-methylation[82, 84, 99], whereas studies on bacterial *C*-MTs from *cis*-AT modules indicate that α-methylation can occur on malonyl-ACP[100, 101]. Recent work successfully excised bacterial *C*-MT domains and demonstrated activity of *C*-MTs from *cis* and *trans*-AT pathways on the β-keto substrates[102, 103]. Still, an even larger mystery surrounding PKS *C*-MTs is their relationship to the ΨMT in mFAS. Understanding how PKS *C*-MTs are accommodated within a module compared to the vestigial mFAS ΨMT could provide hints towards the structure of an ancestral PKS/mFAS, which likely had an active *C*-MT.

PKS O-MT domains

O-MTs encoded in PKS gene clusters can be found as stand-alone proteins, which act as post-PKS tailoring enzymes and selectively methylate hydroxyls on the polyketide after it has been offloaded from the assembly line[19, 104-106]. Alternatively, on-assembly-line *O*-methylation can be catalyzed by *O*-MTs embedded within PKS modules (Figure 1.2). *O*-MTs are typically found in pathways of myxobacterial or cyanobacterial origin from both *cis*-AT[19, 23, 24, 107-113] and *trans*-AT[114, 115] PKS, but are less abundant than *C*-MTs. *O*-MTs are predicted to produce an enol methylether (**6**, Figure 1.4) by methylation of the β-keto intermediate produced by the KS condensation reaction or a β-methoxy (**8**, Figure 1.4) by methylation of the β-hydroxyl product of stereoselective KR reduction. *O*-MTs are members of the class I superfamily but are highly diverged from MTs of known structure and the PKS *C*-MTs. Additionally, *O*-MTs are fully encoded prior to the start of KR domains, suggesting that they are positioned differently than *C*-MTs in the PKS module. Like the *C*-MTs, limited information is available regarding PKS *O*-MT structure and function and little can be predicted regarding their mechanism of catalysis or substrate specificity through bioinformatic approaches[41].

Initiation module MTs

Some PKS loading/initiation modules containing GNAT-like domains include a third type of class I MT (MT_L), which has low sequence identity (<20%) to *C*-MTs or *O*-MTs found in PKS extension modules. The MT_L domain is typically found in between a conserved ~150 amino acid N-terminal region of the loading module, termed the "adaptor region" (AR) and the GNAT-like domain (Figure 1.7). The AR reportedly enhances the *S*-acetyltransfer from CoA to ACP by the GNAT-like domain[31], but has no obvious sequence relationship to entries in the structure database. Some initiation modules lack the AR and contain only an MT_L and GNAT-like domain (Figure 1.7), along with the ACP. The presence of both an AR and MT_L domain in a module can be correlated with the production of a starter unit larger than acetyl (e.g. propionyl, isobutyryl) (Figure 1.7)[21, 45, 97], whereas the inclusion of only an AR or MT_L leads to the production of an acetyl starter unit (Figure 1.7)[31, 93, 116]. As the GNAT-like domain produces acetyl-ACP from malonyl-CoA decarboxylation and *S*-acetyltransfer reactions (Figure 1.3), the MT_L has been proposed to directly methylate acetyl-ACP to the branched chain starter units propionyl[45] and isobutyryl (Figure 1.1, Figure 1.7)[21].

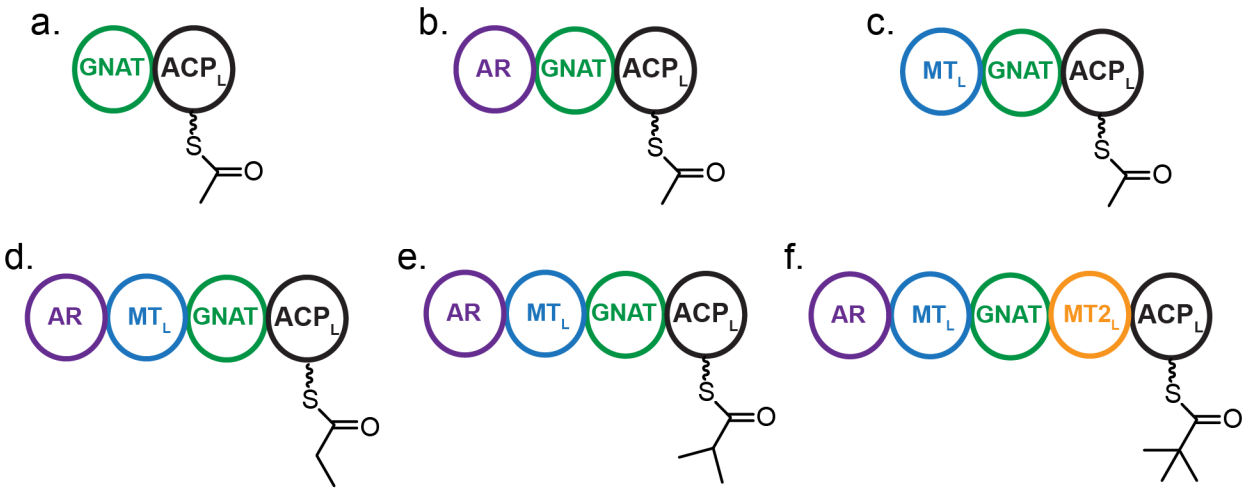


Figure 1.7 The GNAT-like family of initiation modules
PKS initiation modules produce acetyl with modules containing **a.** GNAT-ACP$_L$[94-96], **b.** adaptor region (AR)-GNAT-ACP$_L$[31], and **c.** MT$_L$-GNAT-ACP$_L$[93, 116]. Branched chain starter units **d.** propionyl[45, 97] and **e.** isobutyryl[21] are produced with the inclusion of AR-MT$_L$-GNAT-ACP$_L$ domains. **e.** Production of a *t*-butyl group in the form of pivaloyl-ACP correlates with the addition of a second MT (MT2$_L$) to the initiation module[25].

A fascinating initiation module containing the AR-MT$_L$-GNAT trio of domains followed by a second MT (MT2$_L$) with 30% sequence identity to PKS *C*-MTs and an ACP domain occurs in the apratoxin A[25] (Figure 1.1) biosynthetic pathway. Modules of this type appear to be associated with the formation of *t*-butyl groups (Figure 1.7). The modest sequence identity between PKS C-MTs and MT2$_L$ suggests that MT2$_L$ may differ in function from PKS *C*-MTs and represent a fourth type of PKS MT. MT2$_L$ is likely involved in the formation of the *t*-butyl and may methylate an isobutyryl-ACP substrate produced by the AprA MT$_L$ domain. Additionally, it is worth noting that the GNAT-like enzymes found in the pivaloyl producing modules differ from GNATs in all other initiation modules and lack the conserved His and Thr amino acids required for decarboxylation[31].

Thesis Overview

The following chapters present the full characterization of all four types of MTs found in PKS modules: *C*-MTs, *O*-MTs, MT$_L$, and MT2$_L$, along with an updated analysis of the function of the PKS GNAT-like domains, sparked by discoveries during the characterization of MT$_L$ and MT2$_L$. Experimental determination of the domain boundaries for the *C*-MT encoded in module CurJ from the curacin A biosynthetic pathway (Figure 1.2)[23], enabled direct biochemical

characterization of an isolated *C*-MT from a *cis*-AT PKS and the determination of a crystal structure[117]. Mass spectrometry-based assays clearly demonstrated that α-methylation by the *C*-MT occurs on the β-keto intermediate (**5**, Figure 1.4), after the KS condensation reaction and prior to the action of the reductive modification domains. Catalytic residues, which could not be identified based on sequence information alone, were validated by site-directed mutagenesis leading to a proposed mechanism for the α-methylation reaction. The structure of the CurJ *C*-MT shares many features with the vestigial ΨMT in mFAS, further highlighting the ancestral relationship between PKS and mFAS. Analysis of the sequences bordering PKS *C*-MTs in light of structure permitted modeling of the position of *C*-MTs within the overall PKS module, which resembles the orientation of the ΨMT in mFAS. These insights into PKS *C*-MT structure and function facilitate their use as a biocatalyst in the production of biofuels and commodity chemicals[118].

Similarly, experimental determination of the domain boundaries of PKS *O*-MTs allowed for their excision from the module for biochemical and structural studies. Mass spectrometry-based assays demonstrated that *O*-MTs stereospecifically methylate β-hydroxyl groups. Structures of two *O*-MTs that methylate β-hydroxyl substrates (**7**, Figure 1.4) with opposite stereochemistries from the stigmatellin (Figure 1.1) biosynthetic pathway[19], along with mutagenesis studies, revealed key catalytic residues. Furthermore, the structures of PKS *O*-MTs are distinctly different than PKS *C*-MTs and the mFAS ΨMT, suggesting that the *O*-MTs are a more recent addition to PKS modules. Biochemical and structural characterization of PKS *O*-MTs permits their future development as regioselective biocatalysts and facilitates the future production of *O*-desmethyl polyketides through mutagenesis.

MT_L and $MT2_L$ were examined from the AprA *t*-butyl producing initiation module (Figure 1.7) found in the apratoxin A (Figure 1.1) biosynthetic pathway[25]. Domain boundaries provided from the PKS *C*-MT work[117] and prior work on GNAT-like domains[31] was essential to the design of constructs to probe MT_L and $MT2_L$ activity. A crystal structure of the AprA MT_L-GNAT didomain revealed that the conserved AR is a super-sized lid domain of MT_L and that MT_L is an unusual mononuclear-iron-dependent MT capable of methylating malonyl-ACP to dimethylmalonyl-ACP[119]. Thus, the true MT_L substrate differs from the initially proposed acetyl substrate[45]. The AprA GNAT domain is catalytically inactive, and not capable of acyltransfer or decarboxylation. Thus, the AprA GNAT has been reclassified as a vestigial "pseudo" GNAT

(ΨGNAT). A crystal structure of the AprA ΨGNAT-MT2$_L$ didomain revealed that MT2$_L$ has a nearly identical overall architecture to PKS *C*-MTs, but biochemical assays uncovered unexpected coupled decarboxylation and methylation activities that permit MT2$_L$ to directly transform dimethylmalonyl-ACP to pivaloyl-ACP[120]. As no domain of AprA supported transfer of malonyl from CoA to ACP, the malonyl acyltransferase from fatty acid biosynthesis (FabD), which can malonate AprA ACP$_L$, may be critical for initiating the apratoxin A pathway. Overall, the action of MT$_L$ and MT2$_L$ represents a new mode of generating the unusual *t*-butyl functionality. Furthermore, negative stain EM analysis of the intact AprA module provides insight into loading module architecture and dynamics[120].

The discovery of the identity of AR domains and function of MT$_L$ domains prompted further analysis of the GNAT family of PKS initiation modules with a focus on GphF, the initiation module from the gephyronic acid (Figure 1.1) biosynthetic pathway, that produces an isobutyryl starter unit (Figure 1.7)[21]. The GphF MT$_L$ methylated malonyl-ACP to dimethylmalonyl-ACP via a methylmalonyl-ACP intermediate[119] prior to decarboxylation by the GphF GNAT-like domain. The GphF GNAT-like domain selectively decarboxylated the dimethylmalonyl-product of MT$_L$ to produce the isobutyryl starter unit. A crystal structure of GphF GNAT with bound isobutyryl-CoA facilitates comparison to the GNAT-like enzyme from CurA, which decarboxylates malonyl[31]. Comparative mutagenesis of GphF and CurA GNATs provided additional insight into decarboxylation by GNAT-like domains. Strikingly, GphF GNAT did not display acyltransfer activity, and additional purification of the CurA GNAT markedly reduced the slow acyltransfer activity. We conclude that the primary role of the GNAT-like domain in PKS biosynthesis is decarboxylation, leading to a proposed reclassification of PKS GNAT-like domains as acyl-ACP decarboxylases.

Chapter 2 Domain Organization and Active Site Architecture of a Polyketide Synthase *C*-Methyltransferase

Reproduced with permission from Skiba, M.A., Sikkema, A.P., Fiers, W.D., Gerwick, W.H., Sherman, D.H., Aldrich, C.C., Smith, J.L. Domain Organization and Active Site Architecture of a Polyketide Synthase *C*-methyltransferase. *ACS Chem Biol* 11, 3319-3327 (2016).

Summary

Polyketide metabolites produced by modular type I polyketide synthases (PKS) acquire their chemical diversity through the variety of catalytic domains within modules of the pathway. Methyltransferases are among the least characterized of the catalytic domains common to PKS systems. We determined the domain boundaries and characterized the activity of a PKS *C*-methyltransferase (*C*-MT) from the curacin A biosynthetic pathway. The *C*-MT catalyzes *S*-adenosylmethionine-dependent methyl transfer to the α-position of β-ketoacyl substrates linked to acyl carrier protein (ACP) or a small-molecule analog but does not act on β-hydroxyacyl substrates or malonyl-ACP. Key catalytic residues conserved in both bacterial and fungal PKS *C*-MTs were identified in a 2-Å crystal structure and validated biochemically. Analysis of the structure and the sequences bordering the *C*-MT provides insight into the positioning of this domain within complete PKS modules.

Introduction

Polyketides and their derivatives are among the most chemically and biologically diverse natural products, and their biosynthetic pathways contain a wealth of synthetic potential. Type I polyketide synthases (PKSs) are large multifunctional enzymes that generate complex polyketides with exquisite stereo and regioselectivity [28, 59, 121]. Typical type I PKSs contain acyltransferase (AT) and ketosynthase (KS) domains for the selection of acyl-CoA building blocks and extension of polyketide intermediates as well as an acyl carrier protein (ACP) domain to shuttle intermediates between domains. Additionally, PKSs may have one or more modification domains, such as a

ketoreductase (KR), dehydratase (DH), enoylreductase (ER) and methyltransferase (MT), for further processing of the β-ketoacyl-ACP product of KS extension (Figure 2.1a). In bacteria, type I PKSs are commonly found as modular assembly lines, where each module catalyzes a single two-carbon extension and modification of the polyketide and then passes the intermediate to the next module for further elongation. In some cases, PKS modules lack embedded AT domains and rely on *in trans* ATs encoded in the biosynthetic gene cluster for the delivery of acyl-CoAs. Fungal type I PKSs typically occur as single modules that catalyze multiple rounds of chain elongation and modification [121, 122].

Methyl groups can be introduced at the α-position of a polyketide intermediate through the incorporation of methylmalonyl-CoA extender units selected by the AT or through the action of embedded *C*-methyltransferase (*C*-MT) domains. *S*-adenosylmethionine (SAM)-dependent PKS *C*-MT domains were initially characterized in the lovastatin biosynthetic pathway of the fungus *Aspergillus terreus* and the mixed PKS / non-ribosomal peptide synthetase (NRPS) yersiniabactin pathway [82, 83, 123] and subsequently annotated in multiple fungal PKSs as well as in bacterial *cis*-AT and *trans*-AT pathways [20, 21, 23, 24, 84, 89-92].

PKS *C*-MTs are members of the class I MT superfamily [65, 124], identified by six sequence motifs in the SAM-binding core [125]. Despite structural conservation of the MT core, class I MTs lack conserved active site residues due to the intrinsic reactivity of SAM and the great variety of methyl acceptor substrates. Moreover, numerous loop and domain insertions to the MT core facilitate substrate binding or oligomer formation, resulting in an extremely divergent superfamily with very low overall sequence identity (generally below 20%). *C*-MTs are the only domains commonly found within PKS modules that have remained structurally uncharacterized. Beyond the easily recognized sequence motifs for SAM binding, the PKS *C*-MT domain boundaries are uncertain. However, a pseudo-methyltransferase (ΨMT) domain is present in the PKS homolog, the metazoan fatty acid synthase (mFAS) [1], but the ΨMT lacks the SAM binding motifs and appears to be significantly shorter than the PKS *C*-MTs.

C-MTs could potentially act on the unmodified α-carbon at any point during the PKS catalytic cycle, unlike KS, KR, DH and ER domains, which act in a specific sequence (Figure 2.1b). In an earlier study, it was proposed that PKS *C*-MT domains act after KS condensation and prior to subsequent reductive modifications based on the accumulation of a 2-methyl-1,3-diketo-containing shunt product in a mutagenized lovastatin producing fungus (Figure 2.1b, Route 1) [82].

Methylation of the β-ketoacyl intermediate via Route 1 was observed directly in two fungal PKSs [84, 99]. However, two bacterial PKS *C*-MTs that dimethylate the α-position of polyketide intermediates were shown to act exclusively via Route 2 in epothilone biosynthesis (Figure 2.1b, Route 2) and either on the β-ketoacyl intermediate of Route 1 or on malonyl-ACP preceding KS condensation in yersiniabactin biosynthesis [100, 101]. Thus it was unclear when monomethylation occurs in modular *C*-MTs. Recent work demonstrated *C*-MT activity via Route 1 in bacterial *trans*-AT and *cis*-AT pathways [102, 103].

The biosynthetic pathway for the antimitotic polyketide curacin A (Figure 2.1a) produced by the marine cyanobacterium *Moorea producens* has provided a wealth of structural insight into

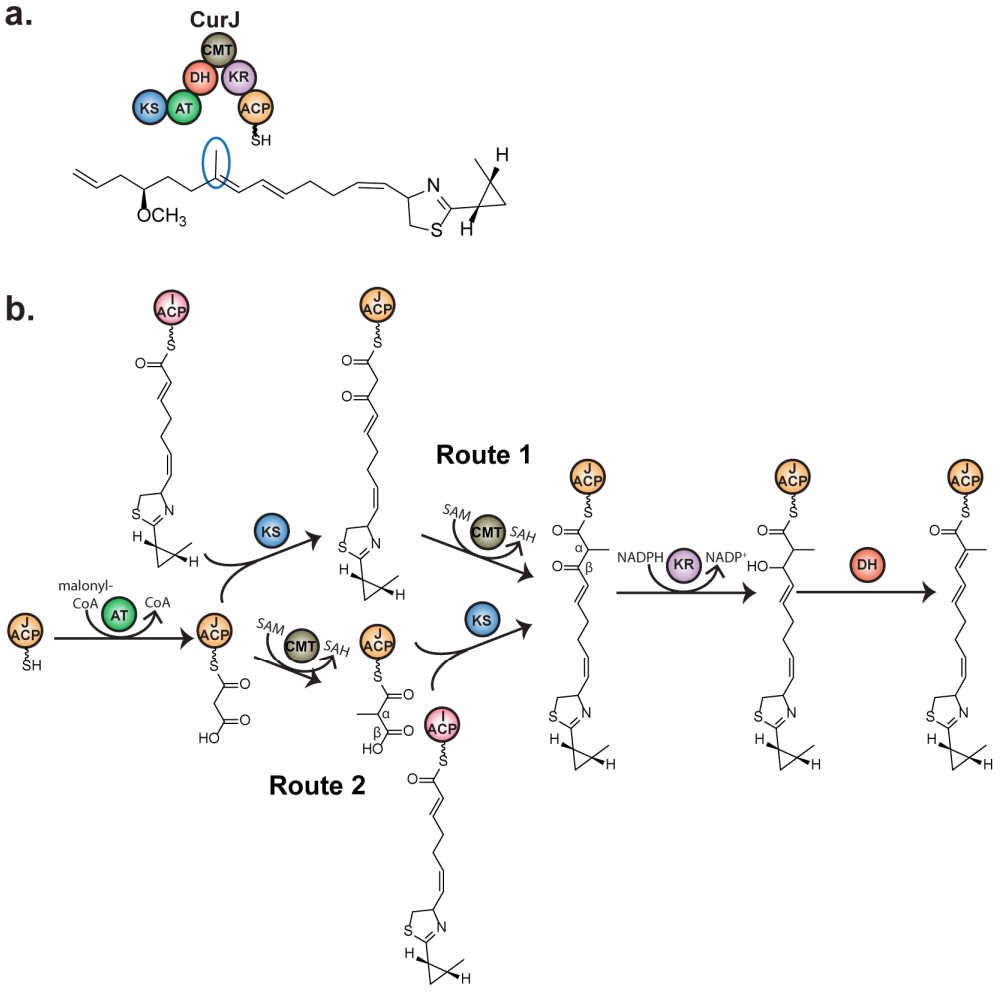


Figure 2.1 Introduction of α-methyl by CurJ *C*-MT.
a. CurJ composition. Curacin A contains a single methyl branch derived from the *C*-MT in module CurJ, which contains ketosynthase (KS), acyltransferase (AT), dehydratase (DH), *C*-methyltransferase (CMT), and ketoreductase (KR) domains. **b.** Potential routes for *C*-methylation in PKSs. Route 1- Methylation occurs on the β- ketoacyl intermediate after the KS condensation reaction. Route 2- Methylation occurs on malonyl-ACP prior to KS condensation.

conventional and unusual PKS biosynthetic capabilities [23, 31, 32, 34, 36, 39, 40, 48, 55, 126, 127]. Modules CurG-CurM catalyze typical PKS extension reactions. Like many PKS pathways from Moorea species, a C-MT domain in CurJ installs a single α-methyl (Figure 2.1a). Here we report the crystal structure of CurJ *C*-MT at 2.1 Å resolution and demonstrate that CurJ *C*-MT follows Route 1 by methylating the α-position of the β-keto acyl substrate. The structure reveals key conserved features of PKS *C*-MT active sites and provides important clues about how *C*-MTs are positioned within PKS modules.

Experimental Procedures

Construct design

All PCR primers are listed in Table 2.1. CurJ C-MT (CurJ residues 1269–1658) was amplified from cosmid pLM9 (GenBank accession code AY652953) [23]. JamJ *C*-MT (residues 1948–2337) and ACP (residues 3166–3302) were amplified from a mixture of jamaicamide A cosmids [24]. *C*-MT and ACP constructs were inserted into pMCSG7 by ligation independent cloning (LIC) to create pMAS179 (CurJ *C*-MT), pMAS195 (JamJ *C*-MT), and pMAS197 (JamJ ACP). The QuikChange protocol (Stratagene) was used for site-directed mutagenesis. DNA sequencing at the UM DNA Sequencing Core verified all constructs and mutations.

Protein expression and purification

Escherichia coli strain BL21(DE3) was transformed with a *C*-MT or ACP plasmid, grown in 0.5 L of TB media with 100 μg/mL ampicillin to an OD_{600} = 1 at 37°C, cooled to 20°C for 1 hr, induced with 200 μM IPTG, and expressed for ~16 hr. Selenomethionine (SeMet) labeled protein was produced in 2 L SelenoMet medium (AthenaES) containing 150 μg/mL of seleno-DL-methionine. Cells were grown to an OD_{600} = 0.6 at 37°C, cooled to 20°C for 1 hr, induced with 200 μM IPTG, and expressed for ~18 hours.

Cell pellets were resuspended in buffer A (50 mM Tris pH 7.4, 300 mM NaCl, 10% (v/v) glycerol, 20 mM imidazole) with 0.1 mg/mL lysozyme, 0.05 mg/mL DNase, and 2 mM $MgCl_2$, incubated on ice for 30 min, lysed by sonication, and cleared by centrifugation (38,760 xg, 30 min). The supernatant was filtered and loaded onto a 5 mL His trap column (GE Healthcare). Proteins were eluted with a 20-400 mM linear gradient of imidazole (buffer B) over 10 column volumes. Gel filtration (HiLoad 16/60 Superdex S200) with buffer C (50 mM Tris pH 7.4, 150 mM NaCl, 10% (v/v) glycerol) was used to further purify proteins. 90 mg/L of *C*-MT was obtained.

Table 2.1 Primers for cloning *C*-MT constructs

CurJ_FM_F	pMAS 35	5'-ATGGAACCGACTACCAAC-3'
CurJ_FM_R		5'-TTATAGAAGTTGATCAAGCTG-3'
CurJ_1269_F	pMAS 179	5'-**TACTTCCAATCCAATGCC**CTACCTCCAGATTTTCTAGATC-3'
CurJ_1658_R		5'-**TTATCCACTTCCAATGCTA**TGTCTCCTTCGTTGGTTCT-3'
JamJ_3166_F	pMAS 196	5'-**TACTTCCAATCCAATGCC**CTGGAGGCAACAGCAC-3'
JamJ_3302_R		5'-**TTATCCACTTCCAATGCTA**TTCATTAATAATTAAATTTAATTCTTG -3'
JamJ_1948_F	pMAS 195	5'-**TACTTCCAATCCAATGCC**CTACCTCCTGATTTCTTGCTG-3'
JamJ_2337_R		5'-**TTATCCACTTCCAATGCTA**TGTCTCCTTCGTTTGTTCTAAG-3'
CurJ_H1548N_F	pMAS 210	5'-GCGGCCAATGTCCTCAATGCAACCACAAGTC-3'
CurJ_H1548N_R		5'-GACTTGTGGTTGCATTGAGGACATTGGCCGC-3'
CurJ_H1548Q_F	pMAS 211	5'-CGGCCAATGTCCTCCAGGCAACCACAAGTC-3'
CurJ_H1548Q_R		5'-GACTTGTGGTTGCCTGGAGGACATTGGCCG-3'
CurJ_H1548A_F	pMAS 209	5'-TATTGCGGCCAATGTCCTCGCTGCAACCACAAGTCTTAAG-3'
CurJ_H1548A_R		5'-CTTAAGACTTGTGGTTGCAGCGAGGACATTGGCCGCAATA-3'
CurJ_E1574A_F	pMAS 212	5'-GGATATTGGTCTTGTATGCAGCAACAACTCGATCTCG-3'
CurJ_E1574A_R		5'-CGAGATCGAGTTGTTGCTGCATACAAGACCAATATCC-3'
CurJ_N1545A_F	pMAS 208	5'-TGATGTAATTATTGCGGCCGCTGTCCTCCATGCAACCACA-3'
CurJ_N1545A_R		5'-TGTGGTTGCATGGAGGACAGCGGCCGCAATAATTACATCA-3'
CurJ_Y1437F_F	pMAS 207	5'-CACAGCAACCCAACTGTTTAAAGACTCAGCTGTAG-3'
CurJ_Y1437F_R		5'-CTACAGCTGAGTCTTTAAACAGTTGGGTTGCTGTG-3'

Bold font indicates handles for ligation-independent insertion into expression vectors

Differential scanning calorimetry (DSC)

Differential Scanning Calorimetry (DSC) was carried out in a Johnson & Johnson ThermoFluor Plate reader in Thermo Scientific black 384 well PCR plates. 0.5 mg/mL of protein in 50 mM Hepes pH 7.4, 150 mM NaCl, 10% (v/v) glycerol, 1 mM CHAPS was mixed with 200 uM 1-anilinonaphthalene-8-sulfonic acid (ANS). Experiments were conducted in triplicate.

Protein crystallization

His-tagged CurJ *C*-MT (native and SeMet) was crystallized by vapor diffusion from 2:2 μL ratio of protein stock (10 mg/mL of CurJ *C*-MT in buffer C with 1 mM SAM or SAH) and well solution (1.30-1.38 M sodium citrate, 4-9 mM glutathione (GSH)/ glutathione disulfide (GSSG), 5% (v/v) acetone) at 20°C. Microseeding was used to obtain single crystals. Thin rod-like crystals

appeared in a few hours and grew overnight. Crystals were harvested directly from the drop in loops and flash cooled in liquid nitrogen.

Data collection and structure determination

Data were collected at GM/CA beamline 23ID-D at the Advanced Photon Source (APS) at Argonne National Laboratory. For the SeMet data set a 10 x 10 μm minibeam was used to collect three wedges of data 0-180°, 60-120°, and 120-180° using inverse beam geometry from three areas on a single crystal due to radiation decay. All data were processed using XDS (Table 2.2, Table 2.3, Table 2.4) [128]. Crystals of wild type *C*-MT grew in stacks and data quality was degraded by streakiness in the diffraction images from even the best single crystals. Single-wavelength anomalous diffraction was used to determine the CurJ *C*-MT structure using ShelXD to locate selenium atoms [129]. SOLVE and RESOLVE were used to determine initial phases, perform density modification (figure of merit = 0.216) and build a partial initial model through AUTOSOL in the Phenix software suite [130, 131]. Phenix AUTOBUILD was used to build 68% of the model. The native crystal form was solved by molecular replacement using a complete model from the SeMet crystal form using Phaser through the Phenix software suite.

Iterative rounds of model building and refinement were carried out to complete the models using Coot and Phenix Refine with eight translation/liberation/screw groups [132]. Electron density is complete for all but the C-terminal nine amino acids of the recombinant protein and includes residues of the TEV sequence. The final structures were validated with MolProbity (Figure 2.2, Figure 2.3, Table 2.2) [133]. Docking of the natural CurJ *C*-MT substrate was performed using Autodock Vina [134] and manually edited. Tunnels were computed using default parameters in CAVER [135]. Figures were prepared in PyMOL [136] and sequence alignments were carried out using Clustal [137] in Jalview [138]. The dendogram was created in Jalview calculated with the average distance using BLOSUM62. The I-TASSER server was used to generate the homology model of the CurJ KR [139-141]. Atomic coordinates and structure factors have been deposited in the Protein Data Bank: 5THY (SeMet CurJ *C*-MT), 5 THZ (wild type CurJ *C*-MT).

Table 2.2 *C*-MT crystallographic information

Data Collection	SeMet	Wild type
Space group	$P2_1$	$P2_1$
Cell Dimensions		
a,b,c (Å)	48.4, 130.4, 62.8	46.1, 129.13, 63.1
α, β, γ (°)	90, 112.0, 90	90, 111.5, 90
X-ray Source	APS 23ID-D	APS 23ID-D
Wavelength (Å)	0.979	1.033
d_{min} (Å)	2.09 (2.16-2.09)[a]	2.10 (2.18-2.10)
R-merge	0.1630 (1.500)	0.1508 (1.048)
Inner-shell R-merge	0.053 (9.3 Å)[b]	0.049 (6.2 Å)[b]
Avg I/σ (I)	8.25 (1.66)	9.3 (1.37)
Completeness (%)	99.6 (99.6)	100 (100)
Multiplicity	10.8 (10.4)	6.6 (6.6)
Total observations	463,157 (44,285)	263,797 (26,230)
$CC_{1/2}$	0.995 (0.409)	0.997 (0.703)
CC*	1 (0.901)	0.999 (0.909)
Refinement		
Data range (Å)	44.48-2.09	43.43-2.10
Reflections used in refinement	42,794	39,893
R_{work}/R_{free} (%)	16.9/21.6	18.5/25.2
Number of non-hydrogen atoms	6355	6220
macromolecules	6049	6040
ligands	40	19
water	328	214
Amino acid residues	765	765
Deviation from ideality		
bond lengths (Å)	0.003	0.007
bond angles (°)	0.60	0.92
Average B-factor	46.3	51.9
protein	46.0	52.0
ligands	86.3	70.3
solvent	46.8	44.5
Ramachandran plot		
favored (%)	98	96
allowed (%)	2.5	3.4
outliers (%)	0	0.4
PDB	5THY	5THZ

[a]Values in parentheses pertain to the outermost shell of data.
[b]d_{min} inner shell

Table 2.3 Scaling statistics for SeMet substituted CurJ *C*-MT (PDB ID 5THY)

Resolution Limit (Å)	Number of Reflections						
	Observed	Unique	Possible	Completeness	R-Factor	I/σ	$CC_{1/2}$
9.33	4751	917	952	96.30%	5.30%	25.74	99.5
6.6	9652	1736	1758	98.70%	6.10%	22.78	99.7
5.39	12627	2212	2214	99.90%	8.20%	19.42	99.5
4.67	13783	2649	2673	99.10%	7.60%	19.69	99.4
4.17	17013	3009	3023	99.50%	8.20%	20.26	99.4
3.81	19730	3351	3360	99.70%	10.30%	18.79	99.4
3.53	19672	3579	3591	99.70%	14.00%	15.83	99.1
3.3	20644	3887	3900	99.70%	16.10%	13.51	98.5
3.11	23017	4159	4160	100.00%	19.90%	11.7	98.3
2.95	24519	4368	4377	99.80%	28.00%	9.22	96.7
2.81	25149	4591	4618	99.40%	34.80%	7.6	95.6
2.69	25745	4818	4815	100.10%	43.70%	6.4	91.8
2.59	26860	5004	5021	99.70%	53.10%	5.31	89.3
2.49	28835	5194	5230	99.30%	65.40%	4.63	85.4
2.41	30358	5356	5357	100.00%	73.20%	4.24	83.1
2.33	31799	5588	5607	99.70%	85.70%	3.61	76.1
2.26	31385	5769	5808	99.30%	98.10%	3.04	65.9
2.2	31616	5905	5900	100.10%	109.40%	2.67	60.5
2.14	34176	6144	6178	99.40%	126.80%	2.3	52.9
2.09	31826	6233	6255	99.60%	150.00%	1.66	40.9
Total	463157	84469	84797	99.60%	16.30%	8.25	99.5

Table 2.4 Scaling statistics for native CurJ *C*-MT (PDB ID 5THZ)

Resolution Limit (Å)	Number of Reflections						
	Observed	Unique	Possible	Completeness	R-Factor	I/σ	$CC_{1/2}$
6.19	9996	1588	1610	98.60%	4.90%	32.67	99.8
4.4	18669	2820	2828	99.70%	6.70%	26.92	99.7
3.6	23050	3598	3613	99.60%	8.90%	22.41	99.6
3.12	29873	4259	4265	99.90%	14.50%	14.74	99.3
2.79	32062	4856	4864	99.80%	25.70%	7.93	98.1
2.55	34617	5289	5303	99.70%	42.20%	4.55	95.2
2.36	39413	5820	5826	99.90%	57.40%	3.04	92.2
2.21	40026	6166	6177	99.80%	84.90%	1.97	94.0
2.08	37616	6194	6615	93.60%	105.10%	1.27	68.4
Total	265322	40590	41101	98.80%	15.10%	9.15	99.7

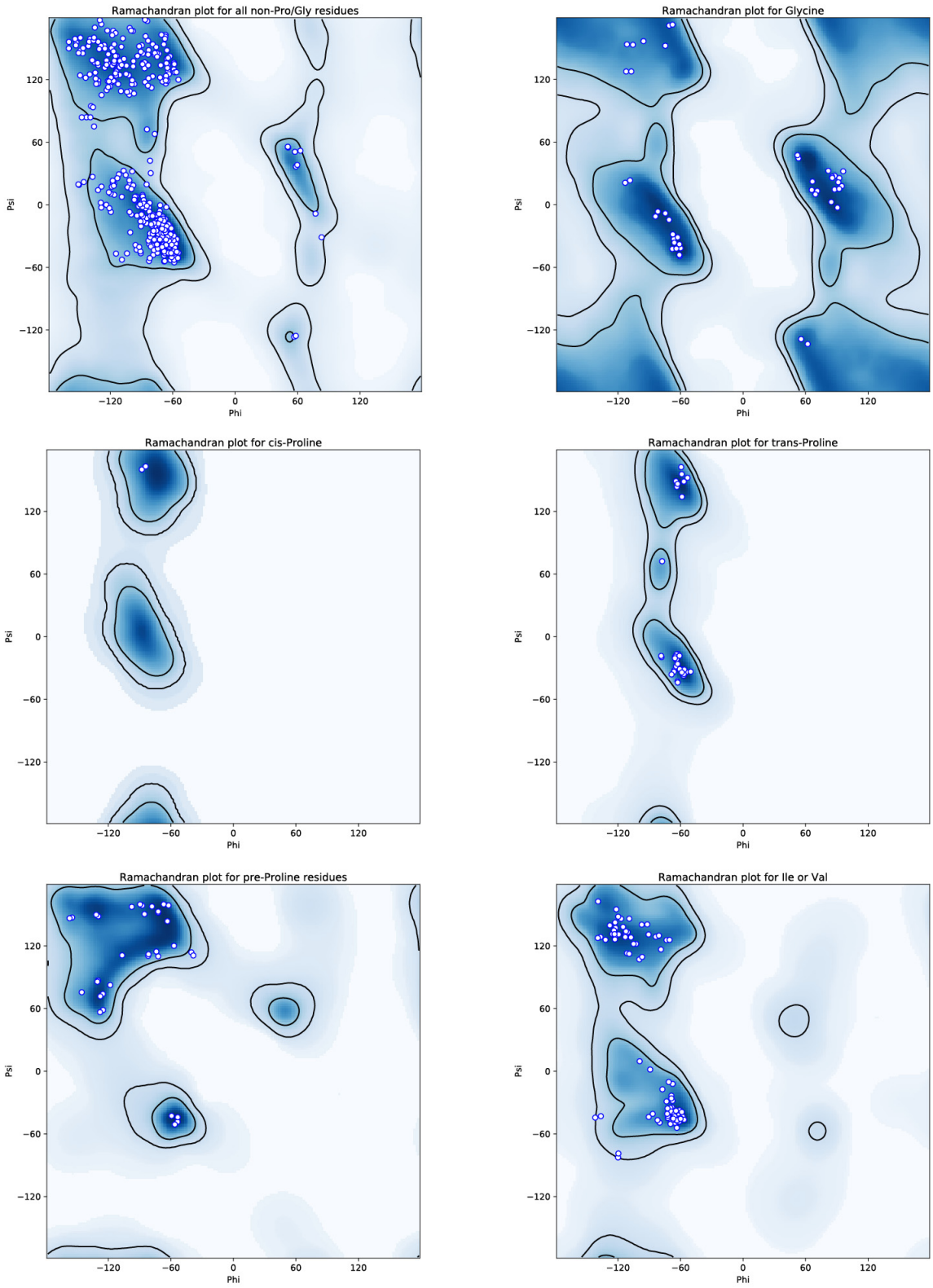


Figure 2.2 Ramachandran plots for CurJ *C*-MT SeMet
Ramachandran plots of final refined model of SeMet labeled CurJ *C*-MT (PDB ID 5THY). Plots were generated using MolProbity[133].

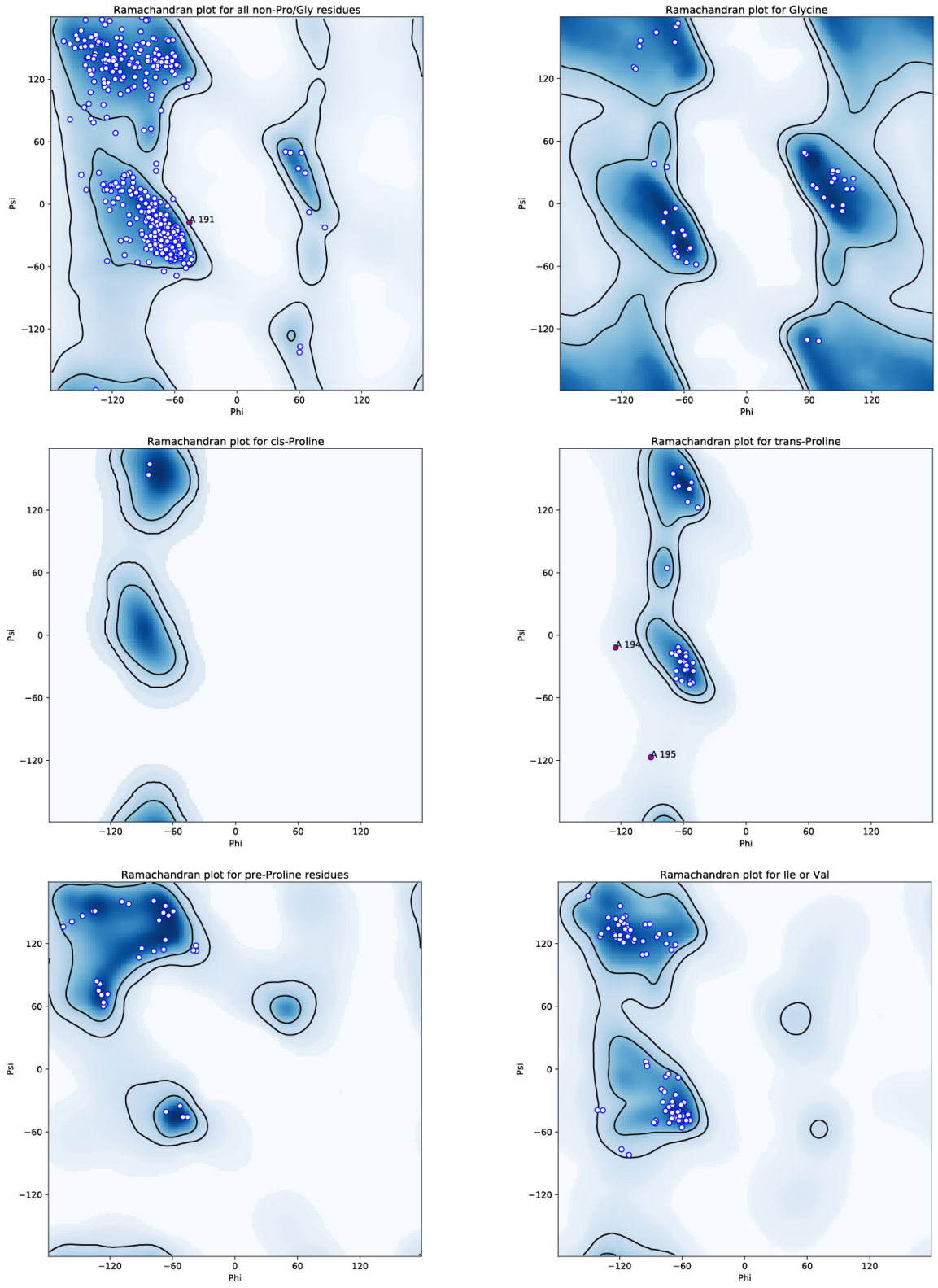


Figure 2.3 Ramachandran plots for native CurJ *C*-MT
Ramachandran plots of final refined model of Native CurJ *C*-MT (PDB ID 5THZ). Plots were generated using MolProbity[133]

Enzyme assays

25 μM enzyme (CurJ *C*-MT, JamJ *C*-MT, or variants) was incubated with 100 μM JamJ ACP, 10 μM *Streptomcyes verticillus* phosphopantethinyl transferase (SVP) [142], 1 mM $MgCl_2$, 1 mM SAM, and 500 μM malonyl-CoA, acetoacetyl-CoA, or DL-β-hydroxybutyryl-CoA in 50 mM HEPES pH 7 (phosphopantetheine ejection) or 7.4 (intact protein), 150 mM NaCl (total volume 50 μL). Reactions were incubated at 30°C for 6 hr (phosphopantetheine ejection) or overnight (intact protein) and quenched with 1% (v/v) formic acid. JamJ ACP from the jamaicamide pathway of another *Moorea producens* strain [24] (82% identical to CurJ ACP) was used as a surrogate due to the toxicity of CurJ ACP to *E. coli* [48]. Nearly identical levels of activity were observed for CurJ and JamJ *C*-MT (Figure 2.6f), indicating that the JamJ ACP is a suitable surrogate for CurJ ACP. NAC-linked substrates (**5**, **7**, Figure 2.6) were synthesized as previously described [143]. For NAC-linked substrate reactions 25 μM *C*-MT was incubated overnight at 25 °C with 1 mM SAM and 500 μM **5** or **7** in 50 mM HEPES 7.4, 150 mM NaCl (50 μL). Reaction mixtures were quenched with 150 μL methanol. Portions of the reaction mixes for ACP-linked substrates (10 μL) and NAC-linked substrates (20 μL) were used for LC-MS analysis (Agilent 6520 Accurate Mass Q-TOF in the positive mode equipped with an Agilent 1290 HPLC system). Separation of samples for intact protein analysis and phosphopantetheine (Ppant) ejection was performed using reverse phase chromatography (Aeris widepore C4 column 3.6 μm, 50 x 2.10 mm) at a flow rate of 0.5 mL/min. Protein was eluted using a gradient of 5-100% B over 4 min (A- H_2O with 0.2% (v/v) formic acid, B- acetonitrile with 0.2% (v/v) formic acid). Intact protein data were analyzed using the Agilent Mass Hunter Qualitative Analysis software with the maximum entropy deconvolution algorithm. Abundance of holo, acetoacetyl, and methylated acetoacetyl phosphopantetheine ejection masses were used to determine percent conversion of CurJ *C*-MT variants [34]. Data were normalized to exclude the spontaneous hydrolysis product (holo-ACP) of acetoacetyl or α-methyl-acetoacetyl-ACP. Three independent reactions were conducted and analyzed for each *C*-MT variant in the phosphopantetheine ejection assay. NAC-linked reaction mixes were separated using a Phenomenex Kinetix reverse-phase C18 column (40 mm x 2.1 mm, 2.6 μm) at a flow rate of 0.5 mL/min. Substrates were eluted using a gradient of 5-100% B over 4 min (A- H_2O with 0.2% (v/v) formic acid, B- acetonitrile with 0.2% (v/v) formic acid).

Results and Discussion

Identification of C-MT domain boundaries

C-MT domains within PKS modules have been annotated through identification of MT motifs (Figure 2.4). However, all have a large unannotated region between the preceding catalytic domain and MT motif I, for example 228 amino acids following the DH domain of CurJ. Moreover *C*-MTs from *cis*-AT and *trans*-AT PKS pathways cluster separately in a phylogenetic tree (Figure 2.5) [21], and have distinct positions within PKS modules, typically preceding KR domains in *cis*-AT pathways and following KR domains in *trans*-AT pathways.

The *C*-MT N-terminal extension is less conserved than the MT core (for example, 30% identity for the CurJ N-terminal extension and 50% for the MT core compared to the homologous regions of the GphI *C*-MT from the myxobacterial gephyronic acid pathway [21]). However, the N-terminal extension is likely to have a defined structure, as it is highly conserved among cyanobacterial *C*-MTs, (for example, 78% identity to CrpA from *Nostoc* sp. ATCC 53780 [89]). We identified approximate domain boundaries based on a multiple sequence alignment of *C*-MTs from cyanobacterial PKS pathways, created fifteen constructs encoding the CurJ *C*-MT with various N- and C-termini (Table 2.5), and tested twelve for production of soluble protein. Ultimately, CurJ residues 1269–1658, herein referred to as 1–390, yielded a pure, stable, recombinant *C*-MT domain

Table 2.5 Solubility of CurJ protein fragments

CurJ Fragment	Protein Solubility
1246-1644	Insoluble
1274-1644	Insoluble
1244-1648	Soluble
1251-1648	No Expression
1257-1648	Soluble
1264-1648	Soluble
1269-1648	Soluble
1244-1658	Soluble
1251-1658	Soluble
1257-1658	Soluble
1264-1658	Soluble
1269-1658	Soluble

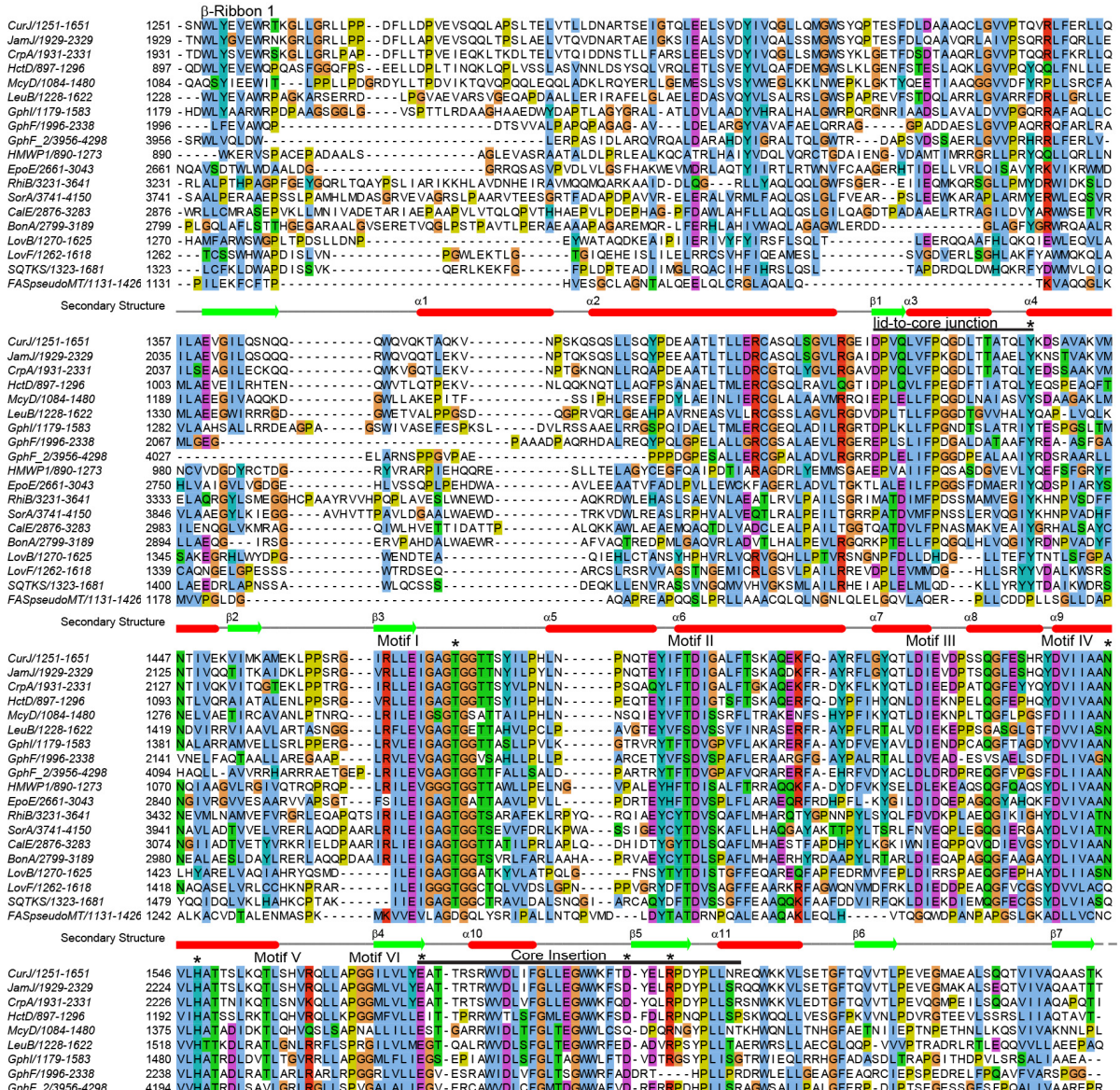


Figure 2.4 Sequence alignment of PKS *C*-MTs with CurJ *C*-MT secondary structure annotation

Key amino acids in or near the active site are indicated by asterisks, including the His280-Glu306 dyad, Tyr169, Thr208, Asn277, Asp329 and Arg333 (His1548-Glu1574, Tyr1437, Thr1476, Asn1545, Asp1597, Arg1601 in CurJ). Amino acid numbers in the figure correspond to the *C*-MT domain within each multi-domain protein. The CurJ *C*-MT sequence numbering the text (1-390) corresponds to CurJ residues 1269-1658. Accession codes for each protein are given in the Figure 2.5 caption.

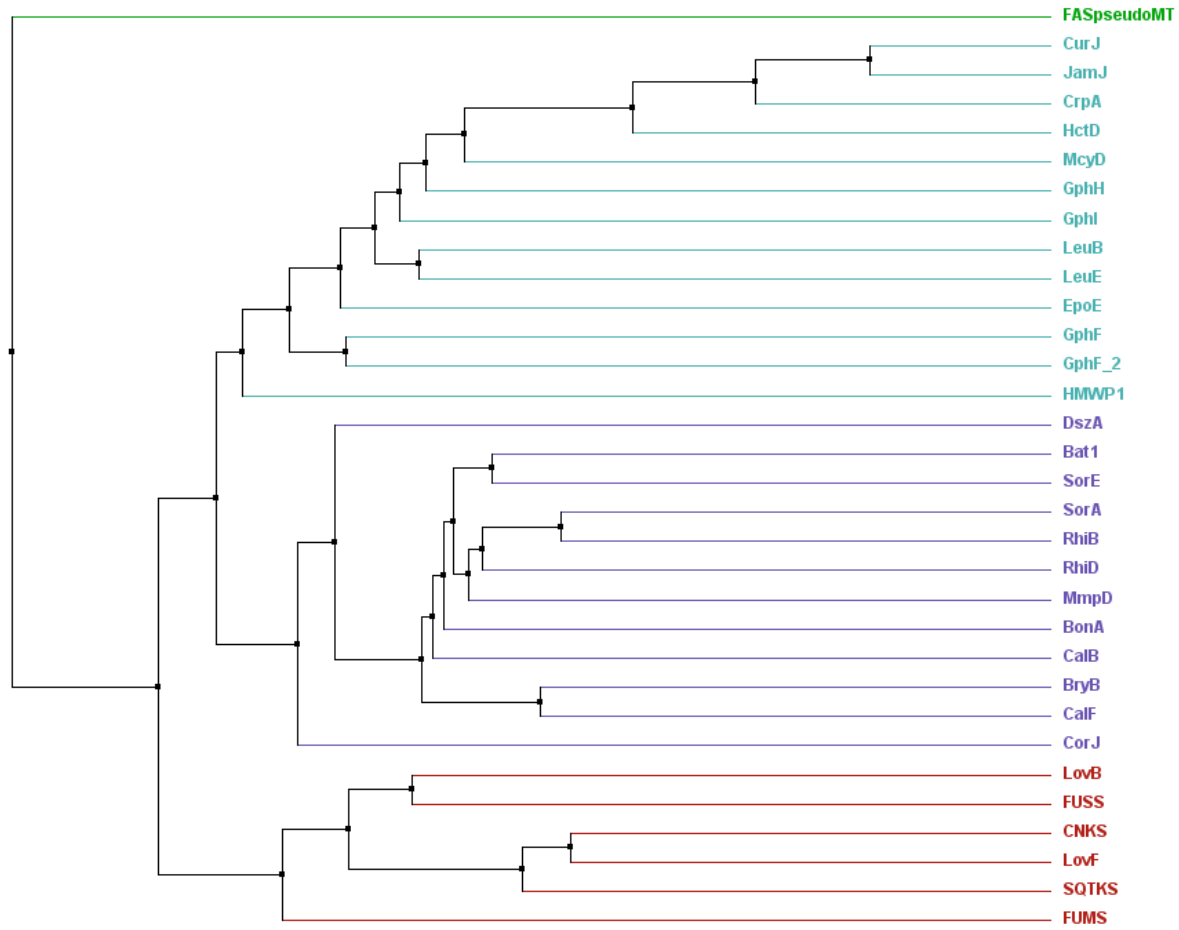


Figure 2.5 Dendogram of PKS *C*-MT domains.
mFAS ΨMT, green; *cis*-AT pathway *C*-MTs, cyan; *trans*-AT pathway *C*-MTs, purple; fungal pathway *C*-MTs, red. Pathway abbreviations (GenBank accession codes) are as follows. FAS ΨMT - *Sus scrofa* (NM_001099930.1), Cur- curacin A (HQ696500.1), Jam- jamaicamide (AY522504.1), Crp- cryptophycin (EF159954.1), Hct- hectochlorin (AY974560.1), Mcy- microcystin (KC699835.1), Gph- gephyronic acid (KF479198.1), Leu- leupyrrin (HM639990.1), Epo- epothilone (AF217189.1), HMWP1- yersiniabactin (AE009952.1), Dsz- disorazole (DQ013294.1), Bat- batumin (GU479979.1), Sor- sorangicin (HM584908.1), Rhi- rhizoxin (AM411073.1), Mmp- mupirocin (AF318063.3), Bon- bongkrekic acid (AFN27480.1), Cal- calyculin (AB933566.1), Bry- bryostatin (EF032014.1), Cor- corallopyronin (HM071004.1), Lov- lovastatin (AF151722.1, AH007774.2), FUSS- fusarin C (AY604568.1), CNKS- compactin (AB072893.1), SQTKS- squalestatin synthase (AY217789.1), FUMS- fumonisin (AY495601.1**).**

that was monomeric in solution as determined by size exclusion chromatography (data not shown).

Methyltransferase activity

Based on precedent for *C*-methylation in PKSs, we expected that CurJ *C*-MT would act on either malonyl-ACP via Route 2 or the β-ketoacyl intermediate formed after KS-condensation via Route 1 (Figure 2.1b) [84, 99, 101-103]. As native CurJ substrates were unavailable, acetoacetyl-ACP (**1**) was tested using an assay based on intact protein mass spectrometry (MS). Complete conversion of acetoacetyl-ACP (**1**) to α-methyl-acetoacetyl-ACP (**2**, Figure 2.6a) was observed by LC-MS. However, no methylation occurred with malonyl-ACP (**3**, Figure 2.6c), establishing that CurJ follows Route 1 and not Route 2.

We also tested a potential methylation Route 3 where the *C*-MT acts after KR reduction of the β-ketoacyl intermediate, *i.e*. on the β-hydroxyacyl intermediate. Reactions performed with DL-β-hydroxybutyryl-ACP (**4**) to mimic the β-hydroxy intermediate failed to generate a methylated product (Figure 2.6e), indicating that the *C*-MT is specific for the β-ketoacyl intermediate. CurJ *C*-MT also methylated an *N*-acetyl cysteamine (NAC-) linked β-ketopentanoyl substrate (**5**, Figure 2.6b) to (**6**), but not β-hydroxypentanoyl-NAC (**7**) (Figure 2.6d).

Table 2.6. Expected and observed mases of ACP species and Ppant ejection fragments

Species	Intact protein calculated mass (Da)	Observed mass (Da)	Ppant ejection calculated mass (Da)	Ppant ejection observed mass (Da)
Apo ACP	18108.3			
Holo ACP	18448.4	18448.9	261.13	261.13
Malonyl ACP	18534.4	18535.4	347.13	347.14
Methylmalonyl ACP	18548.4	None	361.14	None
Acetoacetyl ACP	18532.4	18532.8	345.15	345.16
Methyl-acetoacetyl ACP	18546.4	18547.1	359.16	359.17
D,L- β-hydroxybutyryl ACP	18534.3	18535.5	347.16	347.18
Methyl D,L- β-hydroxybutyryl ACP	18548.5	None	361.18	None

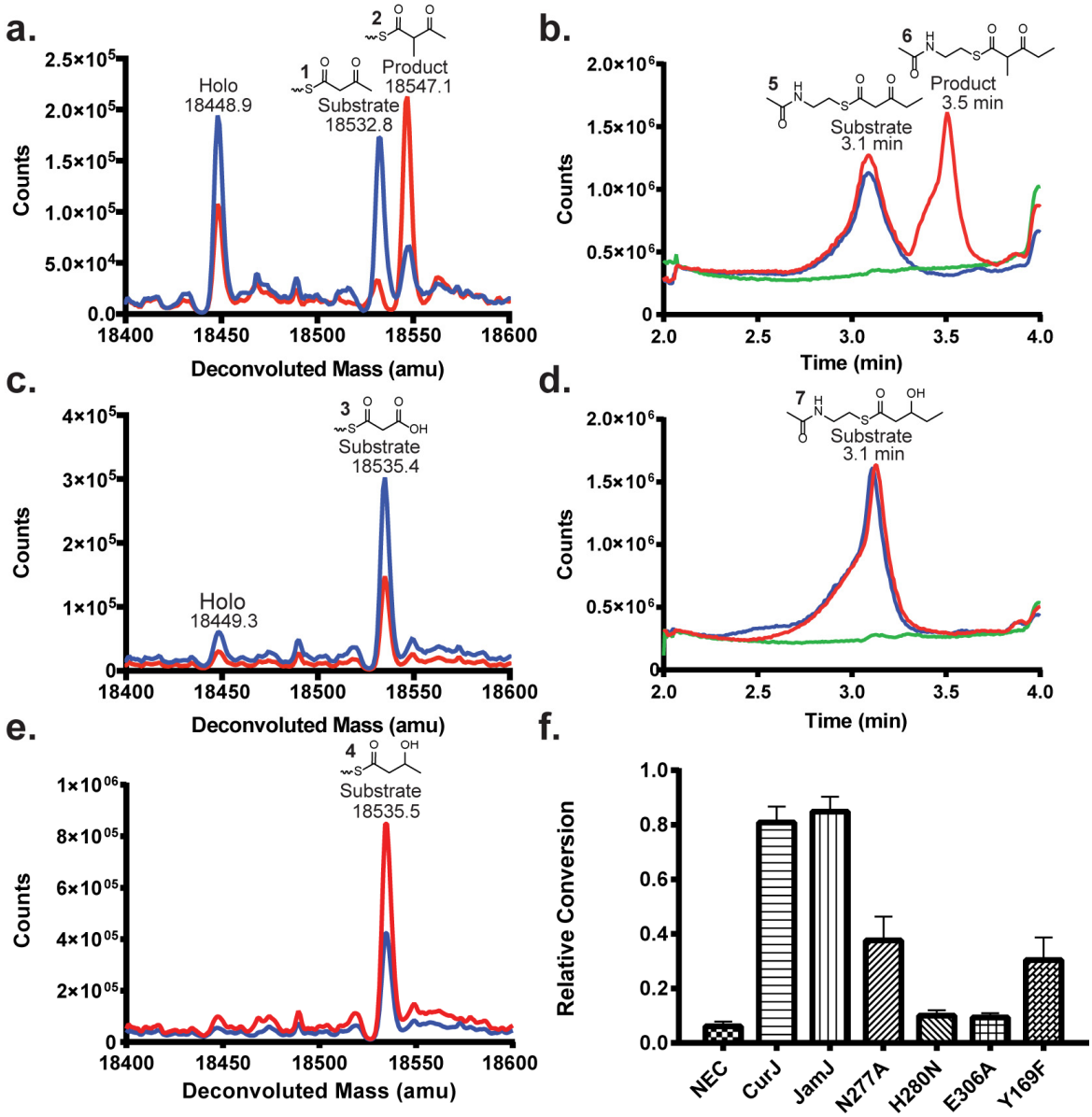


Figure 2.6 LC-MS analysis of *C*-MT activity.
Extracted ion counts are shown in red for reactions via Route 1 in **a.** acetoacetyl-ACP (**1**) and **b.** β-ketopentanoyl-NAC (**5**), via Route 2 in **c.** malonyl-ACP (**3**), and via Route 3 in **d.** β-hydroxypentanoyl-NAC (**7**) and **e.** D,L-β-hydroxybutyryl-ACP (**4**). Blue traces are no-enzyme control reactions; green traces are no-substrate control reactions. See Table 2.6 for expected masses and Figure 2.7 for mass spectra of total ion chromatographs of acyl-NACs. **f.** Relative methylation activities of wild type CurJ *C*-MT, JamJ *C*-MT, and CurJ *C*-MT site-directed mutants with acetoacetyl-ACP (**1**) in an MS assay based on phosphopantetheine ejection (Figure 2.8). NEC = no enzyme control.

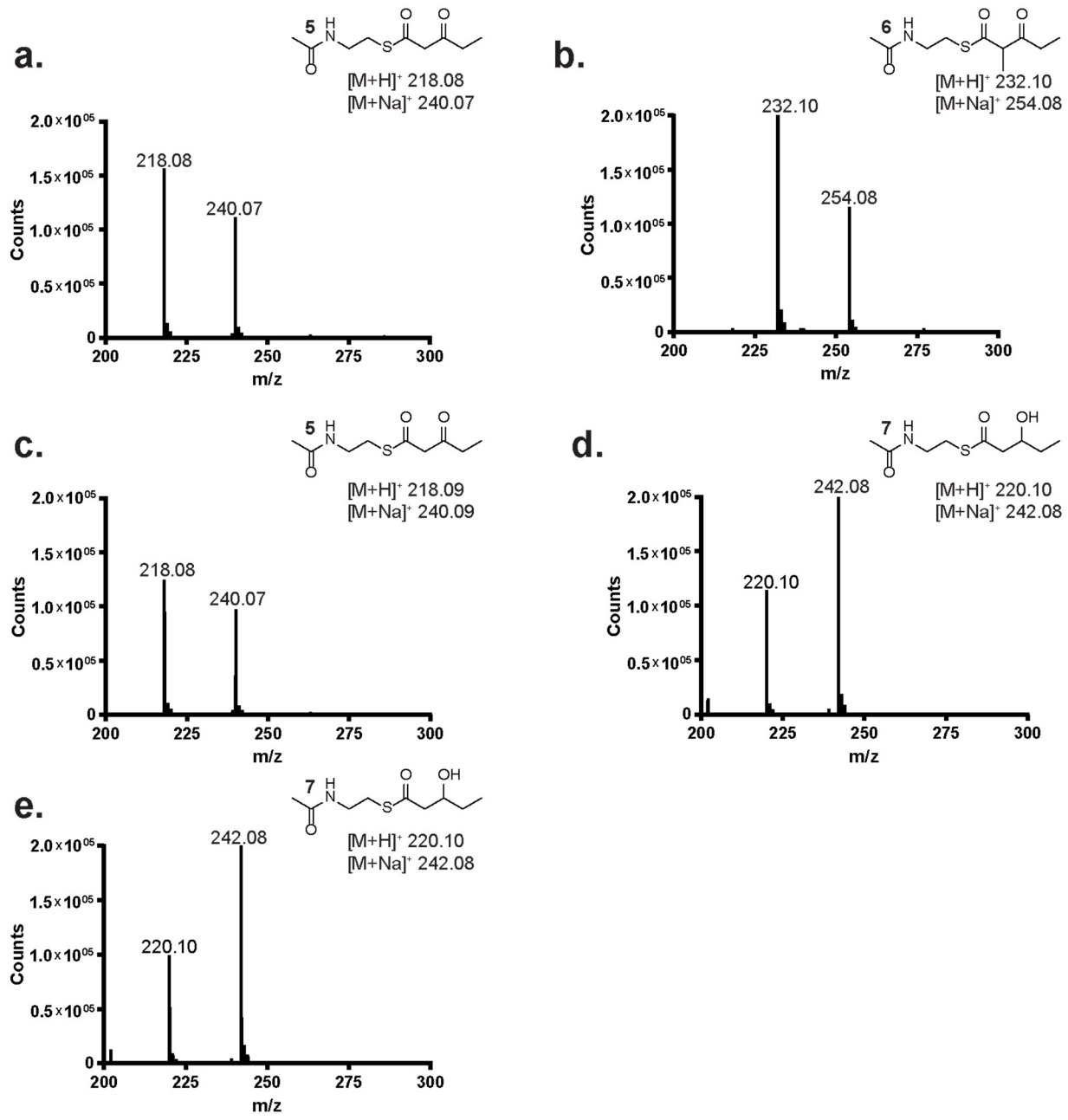


Figure 2.7 Mass spectra of CurJ *C*-MT reactions on NAC substrates
Electrospray-injection (ESI) MS of CurJ *C*-MT reaction mixes for LC peaks in Figs. 2.6b, d. **a.** β-ketopentanoyl-NAC substrate (**5**) peak with LC elution at 3.1 min. **b.** β-ketopentanoyl-NAC product (**6**) peak at 3.5 min. **c.** β-ketopentanoyl-NAC standard (**5**) peak at 3.1 min. **d.** β-hydroxypentanoyl-NAC substrate (**7**) peak at 3.1 min. **e.** β-hydroxypentanoyl-NAC substrate (**7**) standard at 3.1 min. Calculated masses are given beneath each molecular drawing.

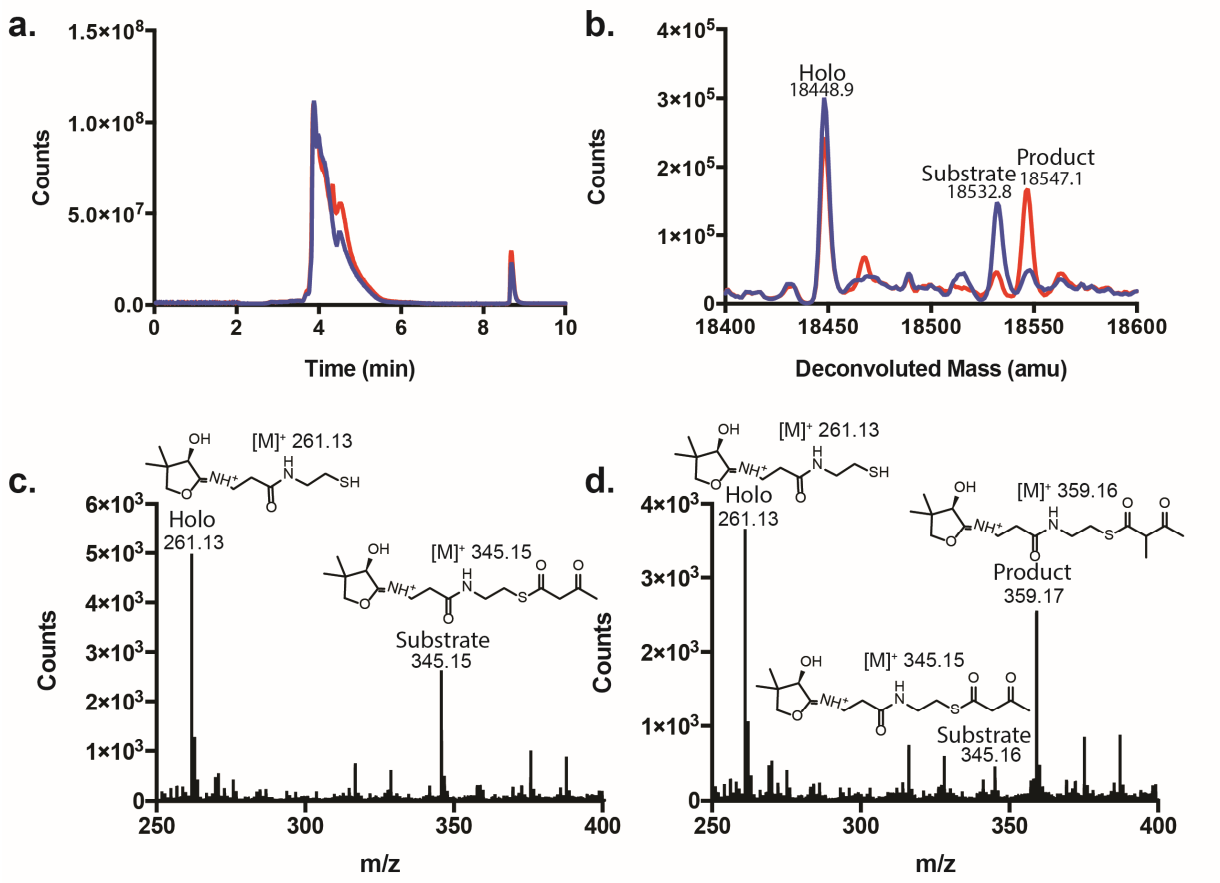


Figure 2.8 Representative data from acetoacetyl-ACP Ppant ejection assay
A phosphopantetheine ejection method was used to detect the acetoacetyl substrate (345.17 Da) and α-methyl-acetoacetyl product (359.17 Da), which were detected and quantified by HR-MS. See Table S1 for observed and expected masses. **a.** Total ion chromatograph for LC trace of reaction mix, red; no-enzyme control, blue. **b.** Deconvoluted ESI-MS for peak eluted between 3.5-4.5 min of ACP-linked substrate and products. A substantial amount of holo-ACP was detected due to the instability of acetoacetyl-ACP. **c.** ESI-MS of phosphopantetheine ejection fragments from the no-enzyme control reaction (blue trace in A and B). **d.** ESI-MS of phosphopantetheine ejection fragments from the *C*-MT reaction (red trace in A and B).

Methylation of the β-ketoacyl intermediate via Route 1 is the most thermodynamically favorable for the formation of the carbanion nucleophile. Among the potential substrates via Routes 1, 2 and 3, the β-ketoacyl intermediate has the lowest pKa ~10 due to the flanking carbonyls. By contrast, the pKa of the α-proton in malonyl-ACP is several units higher as a result of the carboxylate anion. Additionally, the β-hydroxy intermediate produced by the KR has the least acidic α-proton with a pKa estimated ~21, making it the least favorable substrate for methyl transfer.

Structure of CurJ C-MT

The crystal structure of selenomethionine-labeled CurJ *C*-MT was solved to 2.1 Å by single-wavelength anomalous diffraction (SAD) phasing (Figure 2.9 and Table 2.2). The CurJ *C*-MT has a class I MT core domain for SAM binding (residues 1–27, 172–306, 342–381), comprised of a seven-stranded β-sheet flanked by α-helices. A 32-residue insertion (307-341, including helix 13) between β-strand 8 and helix 14 of the MT core distinguishes the PKS *C*-MT from a minimal class I MT (Figure 2.9). The "core insertion" extends over the active site, and is conserved in PKS *C*-MTs from *cis*-AT (49% identity to GphI *C*-MT), *trans*-AT (44% to RhiB) and fungal PKS pathways (44% to SQTKS) (Figure 2.4). The mFAS Ψ-MT has a homologous 26-residue insertion in the equivalent position (Figure 2.10).

Much of the unannotated N-terminal extension of CurJ *C*-MT forms a helical "lid" over the MT core. The large lid (residues 28–171) includes eight helices and a three-stranded β-sheet and is connected to the core by a "lid-to-core junction" containing helices 7 and 8 (Figure 2.9). Helices 1 and 2 (Figure 2.9) form a long amphipathic N-terminal helical "seatbelt" (residues 9–54) that wraps around the MT core and extends into the lid. Nearly all other *C*-MTs from bacterial *cis*-AT pathways contain a lid of similar length with an amphipathic sequence at the N-terminus (Figure 2.4), suggesting that the seatbelt is present in nearly all of them. The seatbelt serves an architectural function of bringing together the N- and C-termini at the bottom of the core domain (15 Å separation). This is analogous to the mFAS Ψ-MT, which lacks helix 1 of the seatbelt and instead contributes a β-strand to the core β-sheet, bringing the domain termini within ~20 Å of one another (Figure 2.10). The mFAS Ψ-MT lid is a truncated variant of the PKS *C*-MT lid.

Knowledge of the *C*-MT domain boundaries within CurJ allowed us to locate a key structure element of the CurJ KR domain (Figure 2.11). An extended β-ribbon is an integral part of PKS KR domains, which are comprised of structural (KR_S) and catalytic (KR_C) sub-domains [2]. The extended β-ribbon contributes to the β-sheets of both KR_S and KR_C, but its two β-strands lie within inter-domain linker sequences, generally flanking the KR_S sub-domain[2-5, 52, 58, 144-147]. However, in CurJ the highly conserved first β-strand of the KR β-ribbon does not flank KR_S, but is located between the DH and *C*-MT domains (amino acids 1254–1261, nine residues following the DH and eight residues preceding the *C*-MT, Figure 2.11). Interestingly, the first β-strand of the mFAS KR β-ribbon is located in the analogous location between DH and Ψ-MT domains.

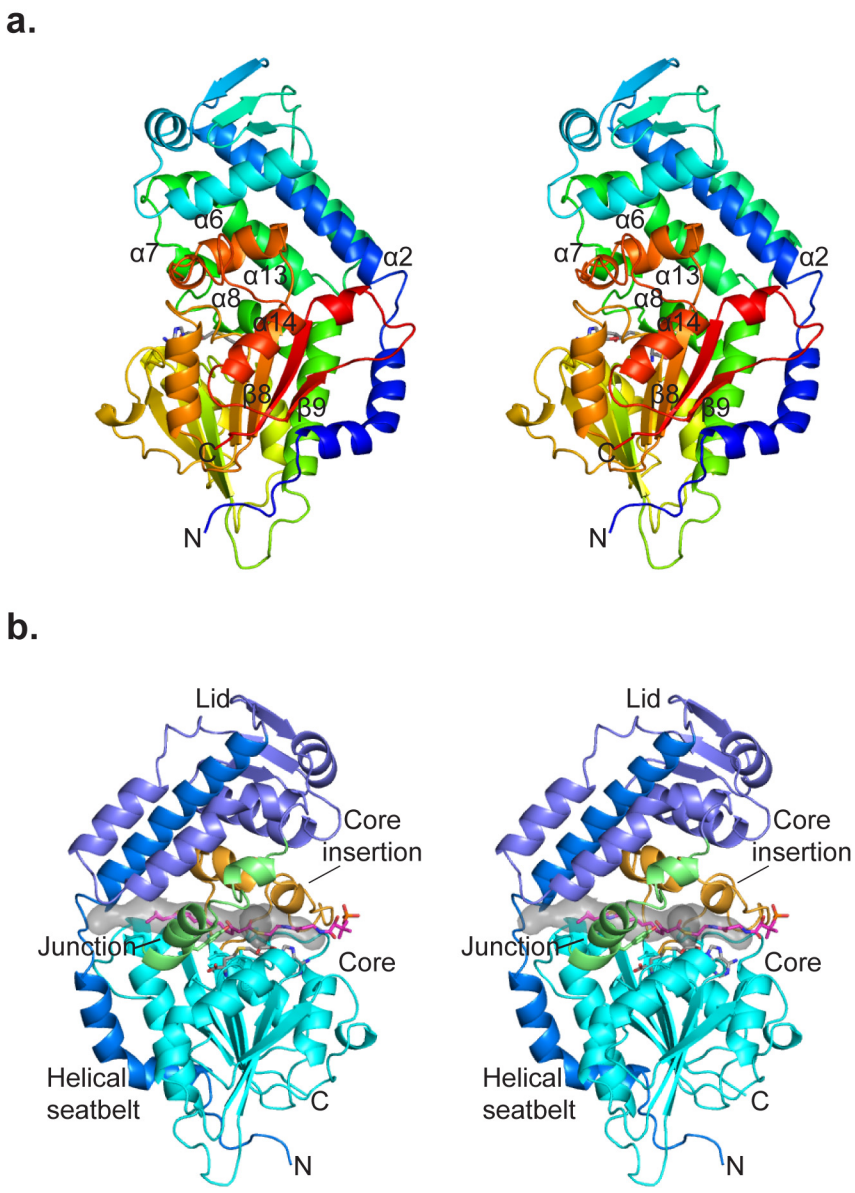


Figure 2.9 CurJ *C*-MT structure.
a. CurJ *C*-MT colored as a rainbow from blue (N-terminus) to red (C-terminus), shown in stereo. SAH is shown in sticks with atomic colors (C, gray; O, red; N, blue; S, yellow). **b.** CurJ *C*-MT colored by structure region. Helical seatbelt, blue; lid, dark blue; lid-to-core junction green; core, cyan; core insertion, orange; SAH, sticks with gray C. The transparent gray surface represents the substrate tunnel, which is lined with hydrophobic residues (Ile35, Phe157, Leu168, Val174, Ala307, Trp313, Val314, Phe318, Leu338). The CurJ *C*-MT substrate and phosphopantetheine (sticks with magenta C) were modeled into the active site. The views in A and B are from opposite sides of the *C*-MT.

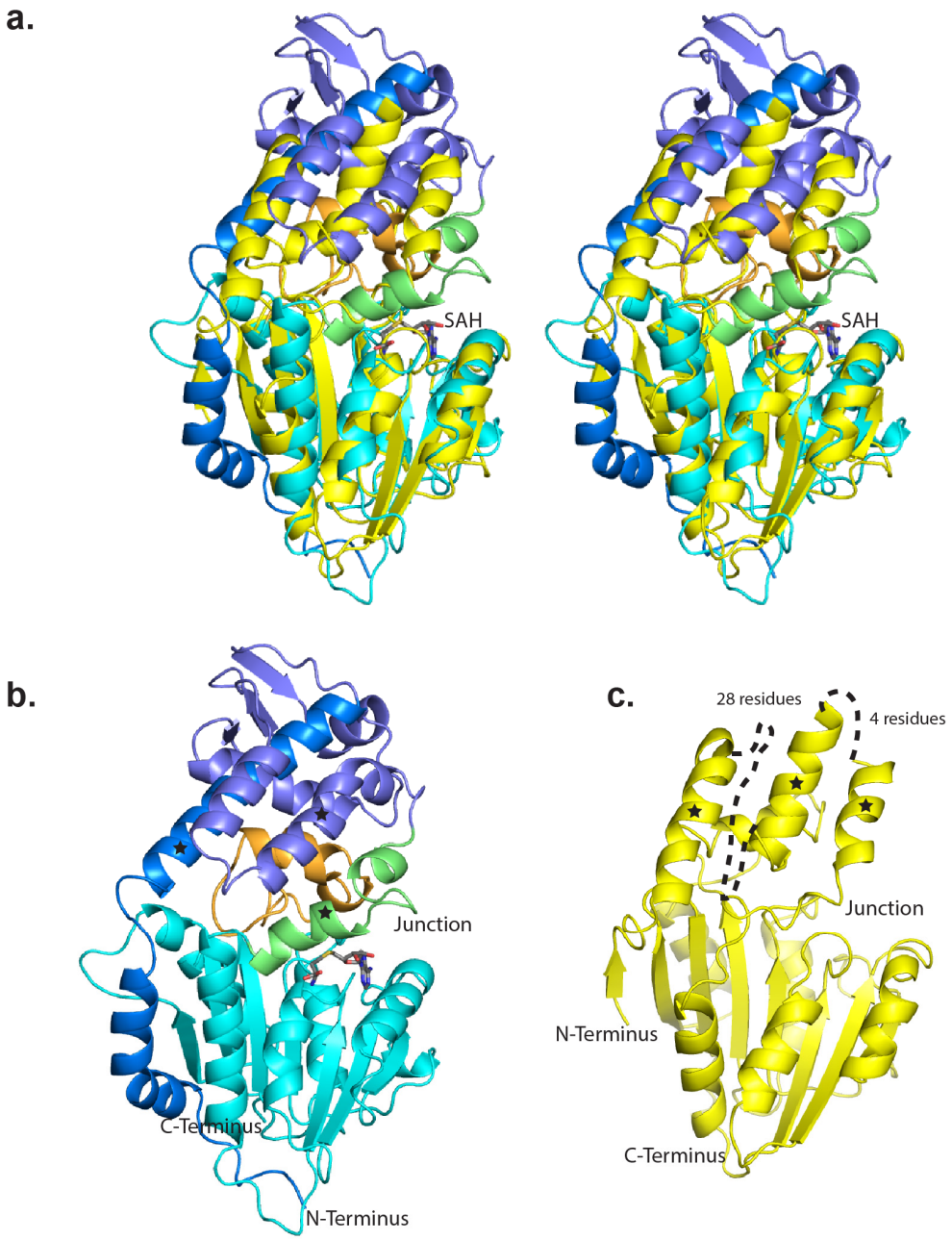


Figure 2.10 Comparison of CurJ *C*-MT and mFAS ΨMT

Comparison of CurJ *C*-MT and mFAS ΨMT. **a.** Overlay of CurJ *C*-MT and mFAS ΨMT superimposed in their MT core domains, in stereo. CurJ *C*-MT is colored in cyan with junction and core insertion colored as in Figure 2.9b. The ΨMT is shown in yellow. SAH is shown in sticks. The CurJ *C*-MT and mFAS superimpose well (RMSD of 1.3 Å for 80 Cα atoms); an insertion to the MT core exists at the same location in both proteins and is shown in orange for the *C*-MT. **b.** CurJ *C*-MT. Lid and junction helices homologous to the mFAS ΨMT lid helices are denoted with stars. **c.** mFAS ΨMT. The structure of mFAS ΨMT lacks many features of the CurJ *C*-MT lid. The remaining lid helices, denoted with stars, are homologous in structure to helices in the CurJ *C*-MT lid and lid-to-core junction.

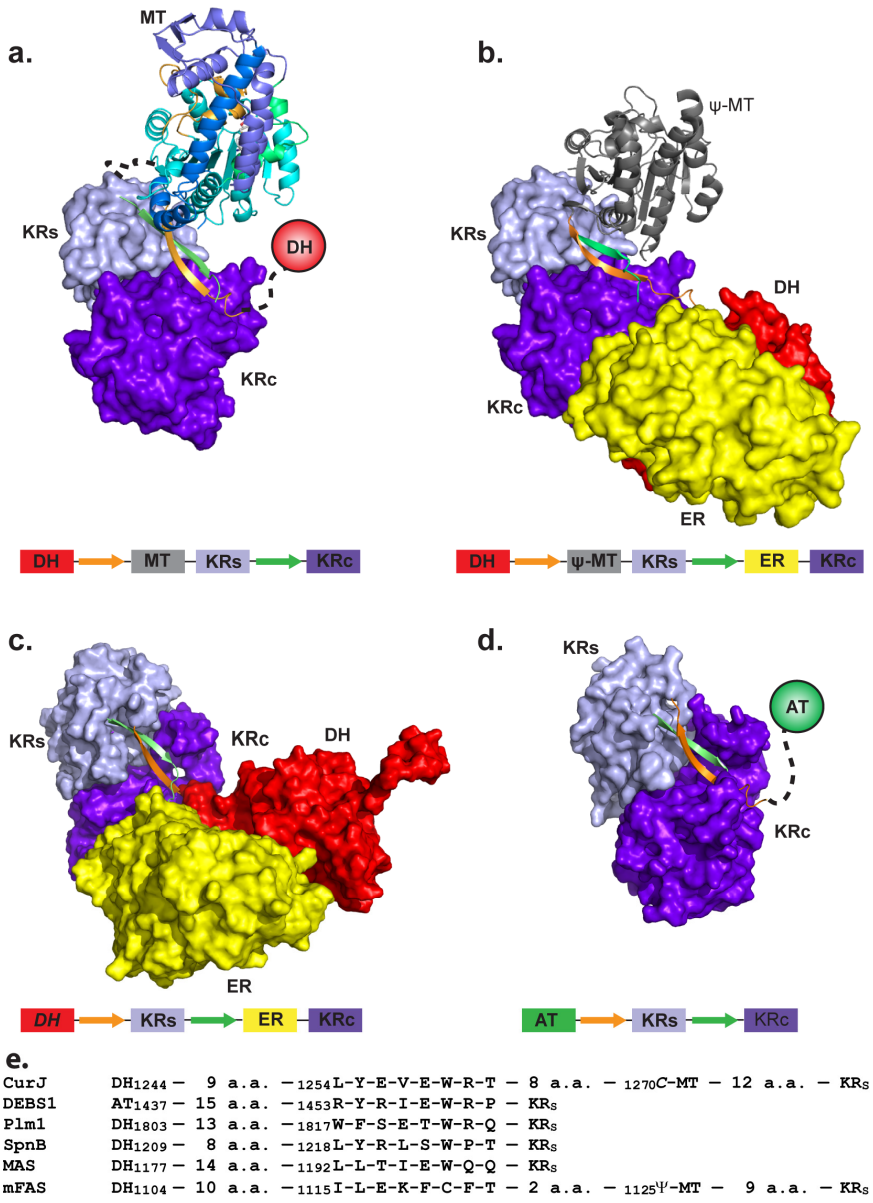


Figure 2.11 Architecture of PKS and mFAS modifying regions.
The β-ribbon that is both an essential part of the KR domain and also an inter-domain linker is orange (1st strand) and green (2nd strand). **a.** Model of CurJ *C*-MT and KR domains, arranged as in mFAS. The model is based on superpositions of the CurJ *C*-MT core with the mFAS ΨMT core and of a PKS KR with the mFAS KR [1, 2]. **b.** mFAS with DH, ΨMT, ER and KR domains [1]. **c.** MAS-like PKS with DH, ER and KR domains [3]. **d.** DEBS1 module 1 modifying region with KR domain only [2]. DH (red), ER (yellow), KRC (purple) and KRS (light blue) domains are represented as surfaces, the CurJ *C*-MT as a ribbon colored as in Fig. 3b, the mFAS ΨMT as a gray ribbon, and other domains as circles. Below each structure is a schematic of domain and β-ribbon strand arrangement colored as in the structure image. **e.** Sequence alignment and domain context of β-ribbon strand 1 in CurJ, DEBS module 1 [2], phoslactomycin module 1 [4], spinosyn module 2 [5], MAS-like PKS [3] and mFAS [1].

Together with the similarity of the *C*-MT and Ψ-MT structures and the similar positions of their N- and C-termini, this remarkable evolutionary conservation is a strong indication that the modifying region of CurJ (DH–*C*-MT–KR–ACP) and other PKS modules is organized similarly to the modifying region of mFAS (DH–Ψ–MT–KR_S–ER–KR_C–ACP), and that the structure of mFAS [1] can guide modeling of the PKS module modifying region (Figure 2.11a).

Active site architecture

A hydrophobic tunnel between the MT core and the lid is long enough to accommodate the full CurJ *C*-MT substrate (Figure 2.9b), consistent with the observed activity on the hydrophobic β-ketoacyl substrate via Route 1, and the lack of activity with the negatively charged malonyl-ACP. *S*-adenosyl-L-homocysteine (SAH) was bound in the MT core below the tunnel, supported by clear electron density (Figure 2.12). The conserved side chains of His280 and Glu306 form a hydrogen-bonded dyad 8 Å from the SAH (Figure 2.13). The His-Glu dyad is invariant in all bacterial and fungal PKS *C*-MTs, thus the imidazole may act as a catalytic base to deprotonate the substrate α-carbon and form a carbanion nucleophile to facilitate the MT reaction. The reactivity of the His-Glu dyad may be enhanced by the surrounding hydrophobic environment (conserved Ala281, Phe318, Trp324, Pro337, Leu338, Trp344). Additionally, conserved Tyr169 is ~4 Å from the presumed SAM sulfonium position, and may facilitate methyl transfer through a CH-O hydrogen bond with the SAM methyl [148]. The invariant Asn277 side chain points into the active site between SAH and the His-Glu dyad. Given its conservation and proximity to the SAM sulfonium, Asn277 could play a role in substrate positioning.

Remarkably, the SAH homocysteine and nearby amino acids are the least well ordered regions of the active site. The two slightly different crystal forms (Table 2.2) provided four independent views of the *C*-MT and revealed small differences in position for the SAH homocysteine, helix 8 in the lid-to-core junction, and MT motif I, which coordinates the SAM/SAH carboxylate (Figure 2.14, Figure 2.12). For two of the four *C*-MT views, density indicated partial disorder for Tyr169 in helix 8 and for the SAH homocysteine, and multiple positions for Thr208 in motif I. In contrast, the SAH adenosine and His-Glu dyad are identically positioned and supported by strong density in all four views. The His280 position is stabilized by two hydrogen bonds within the 277-282 loop and by backbone hydrogen bonds with the conserved Arg333 side chain in the core insertion (Figure 2.13). These observations indicate that the substrate-binding region of the active site is flexible in absence of β-ketoacyl-ACP.

The CurJ *C*-MT core loops, the core insertion, and the lid-to-core junction surround the entrance to the active site, and amino acids in these regions are likely to interact with the ACP (Figure 2.9b, Figure 2.13). An analogous core insertion exists in the only other structurally characterized SAM-dependent enzyme with an ACP-linked substrate, cyclopropane synthase of mycolic acid biosynthesis [149]. As with cyclopropane synthase, the position and the highly conserved sequence of the PKS *C*-MT core insertion are consistent with a role in ACP recognition.

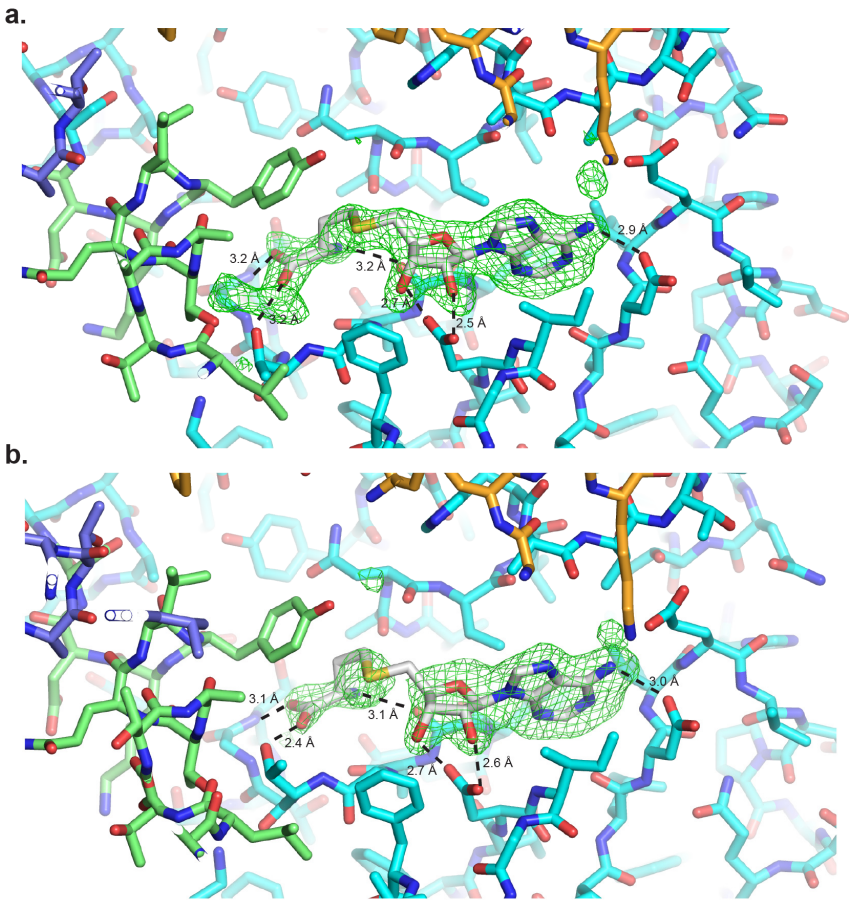


Figure 2.12 Omit density for SAH
Omit density (F_o-F_c, contoured at 3σ) for SAH for the two copies of CurJ *C*-MT in the SeMet crystal form.**a.** chain A, **b.** chain B. The *C*-MT is colored by structure region: dark blue lid, green lid-to-core junction, cyan core, orange core insertion. Hydrogen bonds between SAH and conserved active site residues are shown as dashed lines. Density is excellent for the adenosine moiety of SAH, but the weaker density for the homocysteine portion indicates motion, as do the differences in protein structure shown in Figure 2.14.

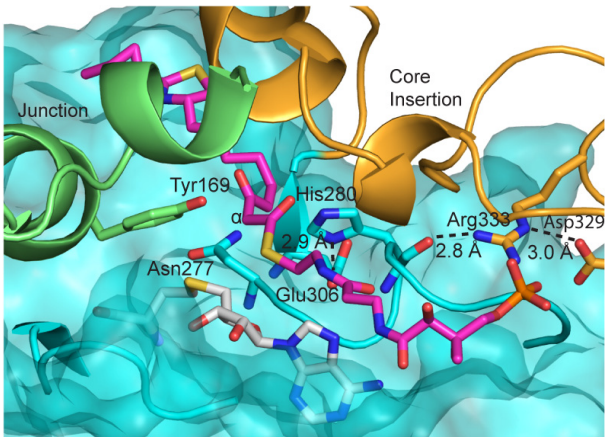


Figure 2.13 CurJ *C*-MT active site.
Key amino acids, modeled substrate (magenta C), and SAH (white C) are shown in sticks. The MT core is represented as a transparent surface with ribbon. The junction and core insertion are shown as ribbons. *C*-MT structural regions are colored as in Figure 2.9b.

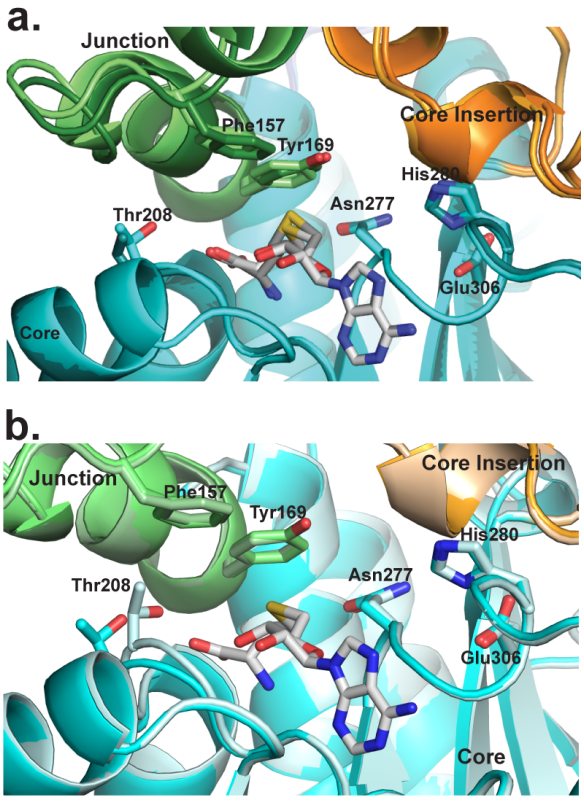


Figure 2.14 Movement surrounding the CurJ *C*-MT active site.
Structural regions are colored as in Figure 2.9b. **a.** Closely related SeMet and wild type CurJ *C*-MT crystal forms (light and dark colors) are different in the lid-to-core junction and SAH homocysteine position. **b.** The position of Thr208 in MT motif I differs in the two *C*-MT chains in the SeMet crystal form.

Activity of CurJ C-MT with active site substitutions

To probe the role of conserved amino acids in the active site, CurJ *C*-MT variants were produced and activity was evaluated using a phosphopantetheine (Ppant) ejection mass spectrometry assay (Figure 2.6f, Figure 2.8) [150]. Substitutions in the active site His-Glu dyad abolished activity, consistent with their proposed role as a catalytic dyad. The melting temperatures of the wild type and the His variants were comparable (Table 2.7), indicating that the reduced activity was not due to a decrease in the structural integrity of the protein. A Y169F variant had three-fold reduced activity, in agreement with a role in facilitating the methyltransfer reaction. The N277A variant reduced activity two-fold suggesting that it plays a role in substrate recognition or positioning in the active site.

Table 2.7 Thermal stability of CurJ *C*-MT variants

Variant	**T_m (°C)**
CurJ *C*-MT WT	57.17 +/- 0.15
H1548N	58.57 +/- 0.93
N1545A	52.50 +/- 0.00
E1574A	54.53 +/- 0.15
Y1437F	57.33 +/- 0.06

Substrate modeling

His208, the strongest candidate for a catalytic base, is 8 Å from the SAM sulfonium on the opposite side of the active site. We modeled the full-length CurJ *C*-MT substrate into the active site tunnel, as attempts to soak small molecule NAC substrates (**5**, **7**, Figure 2.6) yielded additional but uninterpretable density at the presumed substrate binding site between SAH and His280. The full length CurJ *C*-MT substrate was docked between the MT core and lid domain and manually edited to orient the α-carbon in line with the sulfonium (Figure 2.9b, Figure 2.13). While the tunnel can accommodate the full substrate, the sulfonium and His280 were too far apart for both to interact with the α-carbon of the β-ketoacyl intermediate (Figure 2.13). Additionally, hydrogen-bonding partners that could stabilize an enolate intermediate were not apparent. Thus, we propose that β-ketoacyl-ACP binding leads to small movements of the flexible active site elements (lid-to-core junction, core insertion and adjacent loops) and positions the substrate α-carbon, SAM methionine, Tyr169 and enolate-stabilizing hydrogen bonding partners closer to the His-Glu dyad

and in optimal positions for methyl transfer. Due to the open active site in the crystal structures, we are unable to predict the stereochemical outcome of the reaction. Additionally, the acidity of the newly formed stereogenic proton precludes chemoenzymatic assignment. The chiral center in the CurJ *C*-MT product is eliminated by action of the CurJ DH, which catalyzes formation of a vinyl-methyl functionality in the curacin polyketide chain.

Similarities of CurJ C-MT to other PKS C-MTs

The CurJ *C*-MT structure is an excellent model for a pre-catalytic state of *C*-MTs embedded in PKS modules. These PKS MTs share several distinguishing features including the active site His-Glu dyad, the lid-to-core junction helices, and the MT core insertion (Figure 2.4). Overall, *C*-MTs from *trans*-AT pathways and fungal iterative PKSs have lid domains approximately 30 residues shorter than the CurJ *C*-MT and others from *cis*-AT pathways (Figure 2.4, Figure 2.9). Some PKS *C*-MTs produce gem-dimethyl products [20, 101, 103], which the spacious *C*-MT active site could facilitate by substrate re-orientation between methyl-transfer reactions. Although the natural CurJ *C*-MT substrate is quite long, the domain was capable of methylating shorter β-keto acyl substrates, such as acetoacetyl-ACP. This is in contrast to the fungal LovB *C*-MT, which was selective for substrates of similar length to the native substrate and had no activity on an acetoacetyl analog [99]. Increased substrate specificity in iterative fungal *C*-MTs relative to bacterial type I PKS *C*-MTs can be rationalized, as fungal *C*-MTs have evolved to be active in specific rounds of the iterative catalytic cycle, whereas PKS *C*-MTs will encounter only one β-ketoacyl-ACP during the single round of catalysis within the module.

The biochemical and structural characterization of the CurJ *C*-MT is consistent with previous findings that monomethylating PKS *C*-MTs act on β-ketoacyl intermediates following KS mediated chain elongation, and further shows no activity with β-hydroxyacyl intermediates. Additionally, the structure provides new insights regarding the mechanism of α-methylation in both bacterial and fungal PKS *C*-MTs, including identification of a conserved His-Glu dyad and other active site elements that are flexible in absence of the methyl acceptor substrate. Similarities of the CurJ *C*-MT and mFAS Ψ-MT structures suggest that PKS *C*-MTs may share a similar position within a module as the Ψ-MT within the mFAS.

Chapter 3 Structural Basis of Polyketide Synthase *O*-Methylation

This chapter has been submitted: Skiba, M.A.*, Bivins, M.M.*, Bernard, S.M., Fiers, W.D., Kulkarni, S., Wipf, P., Gerwick, W.H., Sherman, D.H., Aldrich, C.C., Smith, J.L. Structural Basis of Polyketide Synthase *O*-Methylation. *ACS Chem Biol.*

*These authors contributed equally

Summary

Modular type I polyketide synthases (PKSs) produce some of the most chemically complex metabolites in nature through a series of multi-enzyme modules. Each module contains a variety of catalytic domains to selectively tailor the growing molecule. PKS *O*-methyltransferases (*O*-MTs) are predicted to methylate β-hydroxyl or β-keto groups but their activity and structure have not been reported. We determined the domain boundaries and characterized the catalytic activity and structure of the StiD and StiE *O*-MTs, which methylate opposite β-hydroxyl stereocenters in the myxobacterial stigmatellin biosynthetic pathway. Like other PKS modification domains, the *O*-MTs displayed substrate stereospecificity. Key catalytic residues were identified in the crystal structures and investigated in StiE *O*-MT via site-directed mutagenesis and further validated with the cyanobacterial CurL *O*-MT from the curacin biosynthetic pathway. Initial structural and biochemical analysis of PKS *O*-MTs supplies a new chemoenzymatic tool, with the unique ability to selectively modify hydroxyl groups during polyketide biosynthesis.

Introduction

Polyketides comprise a wide variety of bioactive natural products, including clinically effective antibiotic[18], antiparasitic[151], immunosuppressant[106], and chemotherapeutic agents[152]. Modular type I polyketide synthases (PKSs) use a series of multienzyme modules to biosynthesize chemically complex polyketides with exquisite stereospecificity and regioselectivity from acyl-coenzyme A (CoA) building blocks.[153] First, the acyltransferase (AT) selects an acyl group (e.g. malonyl, methylmalonyl) from CoA and delivers it to the acyl carrier protein (ACP), which tethers polyketide intermediates via a phosphopantetheine (Ppant) cofactor throughout the enzymatic

assembly line. The acyl group is condensed with the growing polyketide chain by the ketosynthase (KS), elongating the core scaffold of the polyketide by two carbons to produce a β-keto intermediate. The β-keto polyketide can undergo a series of chemical transformations based upon the nature of the modification domains within the module.

Modification domains give rise to the stereochemical complexity and rich chemical diversity of polyketides.[154] *C*-methyltransferases (*C*-MTs) introduce an (*R*)-α-methyl to the β-keto intermediate.[155] Ketoreductases (KRs) stereoselectively reduce the β-keto to a β-hydroxyl group.[156] KR stereoselectivity can be correlated to sequence motifs that identify the KR as A-type or B-type, which produce β-hydroxyl groups in *S*- and *R*-configurations, respectively.[157] The resulting β-hydroxyl can be dehydrated to a *cis*- or *trans*-α,β-double bond by a dehydratase (DH).[158-161] Finally, an enoyl reductase (ER) can catalyze reduction of the α-enoyl polyketide and reintroduce a stereocenter in intermediates with an α-methyl substitutient.[162, 163] The stereospecific and regioselective control provided by the modular nature of PKSs has sparked significant interest in the rational design of PKS pathways for the production of new pharmaceuticals and other high value commodity chemicals.[164]

O-methyltransferases (*O*-MTs) are not canonical PKS domains, but are embedded in PKS modules of several pathways generally of myxobacterial or cyanobacterial origin.[19, 23, 24, 107-115] *O*-MTs are predicted to methylate β-hydroxyl or β-keto groups yielding the respective β-methoxy or enol methylether products (Figure 3.1).[113] Based upon the well established sequence motifs for binding the (*S*)-adenosylmethionine (SAM) methyl donor,[125] the PKS *O*-MTs are members of the large and diverse class I MT superfamily. However, the sequences diverge significantly from MTs of known structure, including the PKS *C*-MTs.[117, 165] The biosynthetic pathway for stigmatellin, an electron transport inhibitor and antifungal agent from the myxobacterium *Stigmatella aurantiaca*, contains two *O*-MTs, which share 40% sequence identity.[19] The structure of the terminal polyketide product and sequence analysis of the KR domains indicate that StiD *O*-MT methylates an (*S*)-β-hydroxyl (Figure 3.1a) produced by the A-type KR encoded in *stiD*, whereas StiE *O*-MT methylates an (*R*)-β-hydroxyl (Figure 3.1b) produced by the StiE B-type KR. As the two *O*-MTs act on opposite stereocenters, we chose the stigmatellin pathway as a model system to investigate PKS *O*-MT structure and function. Biochemical results were further validated with the

CurL *O*-MT from the cyanobacterial curacin biosynthetic pathway[23]. Like StiE *O*-MT, CurL *O*-MT is predicted to methylate an (*R)*-β-hydroxyl group (Figure 3.1b).

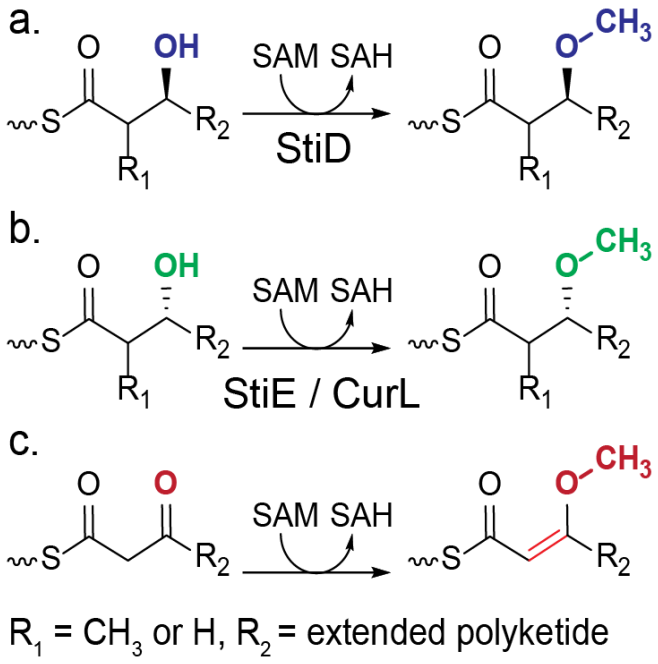


Figure 3.1 Reactions carried out by PKS *O*-MTs
O-MTs methylate **a.** the (*S*)-β-hydroxyl produced by an A-type KR or **b.** the (*R*)-β-hydroxyl produced by a B-type KR, resulting in a β-methoxy group. Some PKS *O*-MTs methylate **c.** a β-keto group, resulting in an enol methylether product. *O*-MTs in this study are indicated below their predicted reactions**.**

Experimental Procedures

Construct design

All primers are listed in Table 3.1 StiD and StiE constructs were amplified from codon optimized DNA (IDT) encoding regions of *stiD* (CAD19088.1) and s*tiE* (CAD19089.1) from *Stigmatella aurantiaca Sg a15*.[19] CurL constructs were amplified from a cosmid library.[23] Constructs encoding StiD *O*-MT (residues 956-1266, pMAS162; residues 976-1266, pMAS165), StiD ACP (residues 1794-1929, pMAS201), StiE *O*-MT (residues 942-1257, pMAS198), StiE ACP (residues 1789-1927, pMAS283), and CurL *O*-MT (residues 981-1315, pMAS411) were inserted into pMCSG7[166] and a construct encoding CurL *O*-MT (residues 951-1315, pMAS/SMB134) was inserted into pMOCR[167] by ligation independent cloning. *stiE* mutations

Table 3.1 Primers for cloning *O*-MT constructs

StiD 976 F	pMAS 163 164 165	**TACTTCCAATCCAATGCC**TCAGCCGTGGATGAAAGC
StiD 956 F	pMAS 160 161 162	**TACTTCCAATCCAATGCC**GAACATCCTGTTGACGGC
StiD 950 F	pMAS 157 158 159	**TACTTCCAATCCAATGCC**GCGAACGGGCAGGGTAAT
StiD 1245 R	pMAS 157 160 163	**TTATCCACTTCCAATGCTA**GGTACTACGAACGTGAGAATC
StiD 1257 R	pMAS 158 161 164	**TTATCCACTTCCAATGCTA**GGCCTCAACCCATTTCTG
StiD 1266 R	pMAS 159 162 165	**TTATCCACTTCCAATGCTA**CATCAGTTCCCGTGCCGC
StiE 961 F	pMAS 200	**TACTTCCAATCCAATGCC**GCGGCGGGGGAAGACG
StiE 951 F	pMAS 199	**TACTTCCAATCCAATGCC**GCTTCGTTCTACGATAGCCTG
StiE 942 F	pMAS 198	**TACTTCCAATCCAATGCC**GCTTCCGAAGGTACATCAGGC
StiE 1257 R	pMAS 198 199 200	**TTATCCACTTCCAATGCTA**CGCAAATTCAGCATACGGTGTC
StiD 1794 F	pMAS 201	**TACTTCCAATCCAATGCC**GCACTGGCCGCCTTAGG
StiD 1929 R		**TTATCCACTTCCAATGCTA**TTTACGACCTTTATTCAGCGCC
StiE 1789 F	pMAS 283	**TACTTCCAATCCAATGCC**GCTCTGCAGTCC
StiE 1927 R		**TTATCCACTTCCAATGCTA**GGCGGTACCTGAG
CurL 951 F	pMAS/ SMB 134	**TACTTCCAATCCAATGCC**TCAAGTTTA
CurL 981 F	pMAS 411	**TACTTCCAATCCAATGCC**CAAAAAATGTTGCCTAAGTTGC
CurL 1315 R	pMAS/ SMB 134 411	**TTATCCACTTCCAATGCTA**AGCTACTTCAGAGTAAGAAGA
StiE Y954F F	pMAS 433	AGGCATCATTGCTTCGTTCTTCGATAGCCTGGTG

StiE Y954F R	pMAS 433	CACCAGGCTATCGAAGAACGAAGCAATGATGCCT
StiE E1102A F	pMAS 432	GATTTAGTGCTCGGATTTGCGGTGGCCGGACTTAT
StiE E1102A R		ATAAGTCCGGCCACCGCAAATCCGAGCACTAAATC
StiE E1102Q F	pMAS 434	TGTATGATTTAGTGCTCGGATTTCAGGTGGCCGGAC
StiE E1102Q R		GTCCGGCCACCTGAAATCCGAGCACTAAATCATACA
StiE L1106H F	pMAS 435	TGAGGTGGCCGGACATATCCCTGACAAGG
StiE L1106H R		CCTTGTCAGGGATATGTCCGGCCACCTCA
StiE Y1209F F	pMAS 437	CAACGTTCCTTTGGCAGCTTTGAGAATGTGTACAAAG
StiE Y1209F R		CTTTGTACACATTCTCAAAGCTGCCAAAGGAACGTTG
StiE Y1223F F	pMAS 439	CGGGGGCCTGATCTCCTTTGTACTGTTTCATG
StiE Y1223F R		CATGAAACAGTACAAAGGAGATCAGGCCCCCG
CurL E1161A F	pMAS 428	CAACCTGGCATTTGGATTTGCAGTAGCTCATCATATTAAGG
CurL E1161A R		CCTTAATATGATGAGCTACTGCAAATCCAAATGCCAGGTTG
CurL E1161Q F	pMAS 430	GATAATTACAACCTGGCATTTGGATTTCAGGTAGCTCATCATATTAAGGAT
CurL E1161Q R		ATCCTTAATATGATGAGCTACCTGAAATCCAAATGCCAGGTTGTAATTATC
CurL Y1281F F	pMAS 429	TGCTGTTAGTAATACAAAGCTAGCCAATCCTTTGCTCAG
CurL Y1281 R		CTGAGCAAAGGATTGGCTAGCTTTGTATTACTAACAGCA
CurL Y1010F F	pMAS 416	ATTTAAAGGTAATGTAGTTTATGACTATTTCAATTCTTTTGCAGAAATTAGTCAAGAAA
CurL Y1010F R		TTTCTTGACTAATTTCTGCAAAAGAATTGAAATAGTCATAAACTACATTACCTTTAAAT
CurL H1165A F	pMAS 413	GGCATTTGGATTTGAAGTAGCTCATGCTATTAAGGATAAATCGCTGTTATTT
CurL H1165A R		AAATAACAGCGATTTATCCTTAATAGCATGAGCTACTTCAAATCCAAATGCC
CurL H1165N F	pMAS 412	GCATTTGGATTTGAAGTAGCTCATAATATTAAGGATAAATCGCTGTTAT
CurL H1165N R		ATAACAGCGATTTATCCTTAATATTATGAGCTACTTCAAATCCAAATGC
CurL Y1267F F	pMAS 408	ATGTTAAGTCAGCTTTTCAATCCTTTAATCAGTTAGGTAAATTACTGAG
CurL Y1267F R		CTCAGTAATTTACCTAACTGATTAAAGGATTGAAAAGCTGACTTAACAT

Primers are listed 5'-3'

Bold text indicates handles for ligation-independent cloning into expression vectors.

were introduced into pMAS198 (E1102A, pMAS432; Y954F, pMAS433; E1102Q, pMAS434; L1106H, pMAS435; Y1209F, pMAS437; Y1223F, pMAS439), *curL* mutations were introduced into pMAS/SMB134 (Y1010F, pMAS416; E1161A, pMAS428; E1161Q, pMAS430; H1165A, pMAS413; H1165N, pMAS412; Y1267F, pMAS408; Y1281F, pMAS429) using the QuickChange protocol (Stratagene). All constructs and mutations were verified by Sanger sequencing at the University of Michigan DNA Sequencing Core.

Protein expression and purification

Plasmids containing *O*-MTs and ACPs were transformed into *Escherichia coli* strain BL21(DE3). Transformed cells were grown in 0.5 L of TB media at 37°C supplemented with 100 μg mL^{-1} ampicillin to an OD_{600}=1-2, cooled to 20°C for 1 hr, and induced with 200 μM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 18 hrs. Media to produce StiD, StiE, and CurL ACP was supplemented with a trace metal mix to insure production of apo-ACP, lacking the phosphopantetheine post translational modification.[168] Selenomethionine (SeMet) labeled StiD *O*-MT (residues 976-1266) was produced in 2L of SelenoMet medium (AthenaES) containing 150 μg/mL seleno-DL-methionine. Cultures were grown to an OD_{600}=0.6 at 37°C, cooled to 20°C for 1 hr, and induced with 200 μM IPTG for 18 hrs.[48]

Cell pellets were resuspended in 35 mL Tris buffer A (50 mM Tris pH 7.4, 300 mM NaCl, 10% glycerol, 20 mM imidazole) with 0.1 mg mL^{-1} lysozyme, 0.05 mg mL^{-1} DNase, and 2 mM $MgCl_2$, incubated on ice for 30 min, lysed by sonication, and cleared by centrifugation (38,760 x g, 30 min, 4°C). The supernatant was filtered and loaded onto a 5 mL His trap column (GE Healthcare). Proteins were eluted with a 5-100% linear gradient of Tris buffer B (50 mM Tris 7.4, 300 mM NaCl, 10% glycerol, 400 mM imidazole) over 10 column volumes. The His-Tag and/or Mocr fusion was removed from StiD *O*-MT (residues 956-1266), StiE *O*-MT (residues 942-1257), CurL *O*-MT (residues 981-1315 and 951-1315), StiE ACP, and CurL ACP via incubation with tobacco etch virus (TEV) protease. The cleavage reaction mixture was dialyzed overnight into Tris buffer A to remove imidazole. Protein lacking the His-tag was isolated by passing over a second His trap column. Proteins were further purified by gel filtration (*O*-MTs, HiLoad 16/60 Superdex S200; ACPs, HiLoad 16/60 Superdex S75) in Tris buffer C (50 mM Tris pH 7.4, 150 mM NaCl, 10% glycerol). Apparent molecular weights were determined by analytical size exclusion chromatography (10/300 Superdex S200 Increase equilibrated with Tris Buffer C).

In order to produce holo StiD ACP, 113 μM StiD ACP was incubated with 0.5 mM coenzyme A (CoA) and 20 μM *Streptomyces verticillus* phosphopantetheinyl transferase (SVP)[142] in Tris buffer C with 20 mM $MgCl_2$ and 2 mM DTT at 20 °C overnight to produce holo-ACP. The His-tag was simultaneously removed by the addition of 12 μM TEV protease. Holo-ACP lacking the His-tag was purified from the reaction mixture by passing over a 1 mL His trap column. Purified holo-ACP was dialyzed into Tris buffer C. StiD ACP concentration was determined using the DC protein assay (BioRad).

Production of acyl-ACPs

N-acetylcysteamine (NAC)-linked diastereomeric triketide substrates (**1**, **2**, **3**, **4**) were synthesized as previously reported[160]. Holo StiD-ACP (50 μM) was incubated with 5 mM **1**, **2**, **3**, or **4** in 300 mM sodium bicarbonate pH 8.1 at 25°C for 4 hrs. ACP was buffer exchanged into reaction buffer (50 mM HEPES 7.4, 150 mM NaCl) and flash frozen. (3*R*)-3-Hydroxy-5-methoxy-myristoyl-CoA was synthesized as previously reported[38]. Apo StiE-ACP (180 μM) was incubated with 36 μM SVP, 0.72 mM (3*R*)-3-hydroxy-5-methoxy-myristoyl-CoA, 20 mM $MgCl_2$ for 1 hr at 25°C in Tris C. The reaction mixture was passed over a 1 mL His Trap column (GE Healthcare) to isolate the ACP. StiE (3*R*)-3-hydroxy-5-methoxy-myristoyl-ACP was further purified via size exclusion chromatography (HiLoad 16/60 Superdex S75) equilibrated with Tris buffer C. (*R*)-3-Hydroxydodecanoyl-CoA was synthesized as previously reported[40]. Apo CurL ACP (180 μM) was incubated with 20 μM SVP, 0.5 mM (*R*)-3-hydroxydodecanoyl-CoA, and 20 mM $MgCl_2$ for 4 hrs at 25°C in Tris buffer C. The reaction mixture was passed over a 1 mL His Trap column (GE Healthcare) to isolate the ACP. CurL (*R*)-3-hydroxydodecanoyl-ACP was further purified via size exclusion chromatography (HiLoad 16/60 Superdex S75) equilibrated with Tris buffer C.

Enzyme assays

StiD triketide-ACP (**1**, **2**, **3**, **4**) (50 μM) was incubated with 25 μM StiD *O*-MT (956-1266) and 0.5 mM SAM in 50 mM HEPES pH 7.4, 150 mM NaCl, 0.5 mM $MgCl_2$. Reaction mixtures (10 μL) were incubated for 24 hrs at 25°C and quenched with 1% formic acid. 0.5 uL of reaction mixtures were subjected to LC/MS analysis. StiE (3*R*)-3-hydroxy-5-methoxy-myristoyl-ACP (**6**) (50 μM) was incubated with 12.5 μM StiE *O*-MT (942-1257) variants or StiE *O*-MT (961-1257) and 0.5 mM SAM in 50 mM HEPES pH 7.4, 150 mM NaCl, 0.5 mM $MgCl_2$. Reaction mixtures (10 μL) were incubated for 15 minutes at 25°C and quenched with 1% formic acid. 1 uL of reaction mixtures were subjected to LC/MS analysis. CurL (*R*)-3-hydroxydodecanoyl-ACP (**7**) (50 μM)

was incubated with 12.5 μM CurL *O*-MT (951-1315) variants and 0.5 mM SAM in 50 mM HEPES pH 7.4, 150 mM NaCl, 0.5 mM $MgCl_2$. Reaction mixtures (10 μL) were incubated for 10 hrs at 25°C and quenched with 1% formic acid. 1 uL of reaction mixtures were subjected to LC/MS analysis. Apo StiD, StiE, or CurL ACP (50 μM) were incubated with 10 uM SVP, 200 μM acetoacetyl-CoA, 0.5 mM SAM and StiD *O*-MT (951-1315), StiE *O*-MT (942-1257), or CurL *O*-MT (951-1315) variants in 50 mM HEPES pH 7.4, 150 mM NaCl, 0.5 mM $MgCl_2$. Reaction mixtures were incubated for 24 hrs at 25°C and quenched with 1% formic acid. 1 uL of reaction mixtures were subjected to LC/MS analysis.

Reaction mixtures were analyzed using the phosphopantetheine (Ppant) ejection method[150, 169] on an Agilent Q-TOF 6545. Samples were separated by reverse phase HPLC (Phenomenex Aeris widepore C4 column 3.6 μM, 50 x 2.10 mm) at a flow rate of 0.5 mL min^{-1} in H_2O with 0.2% (v/v) formic acid. Protein was eluted with a gradient of 5-100% acetonitrile with 0.2% (v/v) formic acid over 4 min. Data were processed using MassHunter Qualitative Analysis software (Agilent). Substrates and products in the StiD *O*-MT reactions experienced in-source decay yielding a conjugated dehydrated species.

Protein crystallization and structure determination

SeMet labeled StiD *O*-MT (residues 976-1266 with His-tag) was crystallized at 4°C by sitting drop vapor diffusion from 2:2 μL mixture of protein stock (10 mg mL^{-1} StiD *O*-MT 976-1266 in Tris buffer C with 1 mM SAM) and well solution (27% PEG 4000, 0.77 M LiCl, 0.1 M Tris pH 7.0). Microseeding was used to obtain single diamond shaped crystals, which grew after 8 days. Crystals were harvested directly from the drop and flash cooled in liquid N_2. Native StiD *O*-MT (residues 976-1266 with His-tag) was crystallized at 20°C by sitting drop vapor diffusion from 1:1 μL mixture of protein stock (11 mg mL^{-1} StiD *O*-MT 976-1266 in Tris buffer C with 1 mM SAM) and well solution (20% PEG 3350, 0.2 M NaF). Diamond shaped crystals grew after one week and were harvested directly from the drop and flash cooled in liquid N_2. StiD *O*-MT (residues 956-1266, His-tag removed) was crystallized at 20°C by sitting drop vapor diffusion from a 2:1 μL mixture of protein stock (13 mg mL^{-1} StiD *O*-MT 956-1266 in Tris buffer C with 1 mM SAM) and well solution (1.5 M ammonium citrate tribasic pH 7.2). Microseeding was used to obtain single rod-shaped crystals, which grew overnight. Crystals were cryoprotected with well solution supplemented with 20% glycerol and flash cooled in liquid N_2. StiE *O*-MT (961-1257 with His-tag), was crystallized at 4°C by sitting drop vapor diffusion from 2:1 μL protein stock

(11 mg/mL His-tagged StiE *O*-MT in Buffer C with 1 mM SAM or SAH) to well solution (10% PEG 3350, 0.1 M sodium formate). Microseeding was used to obtain single crystals. Square bipyramidal crystals grew in 3 days. Crystals were harvested directly from the drop and flash cooled in liquid nitrogen. StiE *O*-MT (residues 942-1257, His-tag removed) was crystallized at 4°C by sitting drop vapor diffusion from a 2:1 µL protein stock (9 mg mL^{-1} StiE *O*-MT 961-1257 in Tris buffer C with 1 mM SAM) and well solution (25% PEG 8000, 0.1 M HEPES pH 7.4) at 4°C. Square bipyramidal crystals grew after 5 days. Crystals were cryoprotected in well solution supplemented with 20% glycerol and flash cooled in liquid N_2.

Diffraction data for all structures were collected at APS beamline 23ID-D or ID-B and processed using XDS (Tables 3.2-3.7).[128] The SeMet StiD *O*-MT 976-1266 was solved by single-wavelength anomalous diffraction (SAD) phasing using Phenix AutoSol[170] in the Phenix Software suite.[130] A nearly complete model was build using Phenix AutoBuild.[171] Native StiD *O*-MT 976-1266 was isomorphous with the SeMet crystal form. StiD *O*-MT 956-1266 and StiE *O*-MT 942-1257 were solved by molecular replacement using Phaser[172]. Iterative rounds of model building and refinement were carried out using Coot[132] and Phenix.refine[173] with automated translation/liberation/screw group selection. Structures were validated using MolProbity (Figures 3.2-3.6).[133] Homologs in the structure database were identified using the DALI server.[174] Sequence alignments were prepared using Clustal[137] through Jalview[138] and figures were prepared with PyMol.[136]

Table 3.2 *O*-MT crystallographic information

Protein	SeMet StiD 976-1266	StiD 976-1266	StiD 956-1266	StiE 961-1257	StiE 942-1257
Ligand	SAH	SAH	SAH	SAM	SAM
Data Collection					
Space group	*P*1	*P*1	$P4_1$	$P4_32_12$	$P4_32_12$
Unit cell, a,b,c (Å)	38.3, 52.4, 68.7	40.2, 56.2, 72.3	71.1, 71.1, 123.0	90.9, 90.9, 84.9	88.6, 88.5, 86.1
α,β,γ (°)	89.6, 88.3, 81.1	86.3, 84.9, 75.8	90, 90, 90	90, 90, 90	90, 90, 90
X-ray source	APS 23ID-D	APS 23ID-D	APS 23ID-B	APS 23ID-D	APS 23ID-D
Wavelength (Å)	0.979	1.033	1.033	1.033	1.033
d_{min} (Å)	1.96 (2.03-1.96)	1.80 (1.86-1.80)	1.70 (1.76-1.70)	1.42 (1.47-1.42)	1.90 (1.99-1.90)
R-merge	0.066 (0.82)	0.073 (0.74)	0.084 (2.34)	0.064 (1.03)	0.068 (2.28)
Avg I/σ(I)	8.7 (1.1)	9.9 (1.6)	19.8 (1.1)	29.7 (2.7)	21.2 (1.1)
Completeness (%)	96.7 (87.0)	96.7 (94.0)	99.7 (99.9)	99.2 (93.7)	100 (99.7)
Multiplicity	3.5 (2.7)	3.5 (3.5)	13.8 (13.8)	24.3 (19.9)	13.0 (13.1)
Total observations	127,763 (8,823)	193,662 (18,628)	931,727 (92,530)	1,638,916 (124,591)	358,381 (35,303)
Wilson B factor (Å^2)	38.2	27.9	31.3	18.2	41.9
$CC_{1/2}$	0.998 (0.559)	0.997 (0.597)	1.00 (0.442)	1.00 (0.803)	1.00 (0.482)
CC*	1.00 (0.847)	0.999 (0.865)	1.00 (0.783)	1.00 (0.944)	1.00 (0.806)
Refinement					
Data range (Å)	41.3-1.96	38.8-1.80	46.5-1.70	40.6-1.42	44.3-1.90
Reflections	36,668	54,837	67,382	67,269	27,550
R_{work}/R_{free} (%)	17.9/22.7	16.8/19.4	16.8/21.2	16.7/18.5	18.3/22.6
Non-hydrogen atoms (#)	4,629	4,995	4,870	2,768	2,480
protein	4,461	4,628	4,506	2,462	2,311
ligands	52	52	52	27	27
water	116	314	312	315	118
Amino acid residues	554	564	561	279	282
Deviation from ideality					
bond lengths (Å)	0.004	0.007	0.014	0.005	0.007
bond angles (°)	0.99	1.15	1.24	0.88	0.84
Average B-factor (Å^2)	54.6	42.4	48.0	30.5	69.9
protein	54.7	42.1	47.6	29.5	70.3
ligands	54.3	53.9	74.2	18.7	58.1
solvent	51.0	45.7	50.1	39.2	60.9
Ramachandran plot					
favored (%)	97.6	97.7	97.8	97.8	97.1
allowed (%)	2.4	2.3	2	2.2	2.9
outliers (%)	0	0	0.2	0	0
PDB ID	6ECU	6ECV	6ECW	6ECT	6ECX

[1]values in parentheses designate outer shell

Table 3.3 Scaling statistics for SeMet StiD 976-1266 (PDB ID 6ECU)

Resolution Limit (Å)	Number of Reflections						
	Observed	Unique	Possible	Completeness	R-Factor	I/σ	CC 1/2
8.77	1599	417	421	99.00%	3.80%	24	99.8
6.2	2736	765	773	99.00%	4.60%	21.63	99.6
5.06	3304	990	1005	98.50%	4.70%	20.18	99.5
4.38	4262	1174	1192	98.50%	3.50%	21.41	99.7
3.92	4959	1330	1351	98.40%	3.80%	21.02	99.7
3.58	5294	1457	1481	98.40%	4.30%	19.21	99.7
3.31	5463	1576	1604	98.30%	5.30%	16.42	99.6
3.1	5728	1736	1762	98.50%	7.00%	13.76	99.2
2.92	6426	1804	1834	98.40%	8.30%	12.15	99.1
2.77	7001	1922	1952	98.50%	9.60%	10.88	98.8
2.64	7604	2052	2090	98.20%	11.00%	9.36	98.8
2.53	7813	2090	2132	98.00%	15.10%	7.38	97.8
2.43	7964	2218	2264	98.00%	18.20%	6.24	96.5
2.34	7930	2283	2338	97.60%	23.00%	5.03	94.9
2.26	7893	2328	2390	97.40%	26.80%	4.32	93.0
2.19	8526	2464	2542	96.90%	34.50%	3.55	90.3
2.13	8901	2467	2546	96.90%	42.50%	2.86	88.2
2.07	9314	2588	2673	96.80%	55.10%	2.28	81.4
2.01	9600	2619	2709	96.70%	71.20%	1.82	70.8
1.96	5448	2396	2850	84.10%	87.20%	0.94	48.7
Total	127765	36676	37909	96.70%	6.60%	8.7	99.8

Table 3.4 Scaling statistics for StiD 976-1266 (PDB ID 6ECV)

Resolution Limit (Å)	Number of Reflections						
	Observed	Unique	Possible	Completeness	R-Factor	I/σ	CC 1/2
5.36	7196	2102	2160	97.30%	5.00%	24.56	99.6
3.8	13335	3756	3834	98.00%	4.90%	24.11	99.5
3.11	16354	4834	4980	97.10%	5.40%	19.92	99.4
2.69	21081	5752	5875	97.90%	6.80%	15.6	99.2
2.41	23090	6520	6681	97.60%	9.40%	10.93	98.7
2.2	25453	7166	7378	97.10%	13.30%	8	98.0
2.04	28579	7767	8038	96.60%	21.50%	5.67	96.0
1.91	28426	8246	8598	95.90%	38.70%	3.3	85.7
1.8	30157	8699	9206	94.50%	66.70%	1.82	66.2
Total	193671	54842	56750	96.60%	7.30%	9.92	99.6

Table 3.5 Scaling statistics for StiD 956-1266 (PDB ID 6ECW)

Resolution Limit (Å)	Number of Reflections						
	Observed	Unique	Possible	Completeness	R-Factor	I/σ	$CC_{1/2}$
5.06	35018	2605	2614	99.70%	3.10%	74.59	100.0
3.59	61842	4612	4612	100.00%	3.70%	69.72	100.0
2.93	86526	5962	5962	100.00%	5.50%	49.69	99.9
2.54	97644	6992	6993	100.00%	9.90%	28.98	99.8
2.27	108050	7944	7944	100.00%	18.80%	17.1	99.5
2.08	124651	8791	8792	100.00%	32.10%	10.72	98.6
1.92	132926	9502	9502	100.00%	58.10%	5.46	94.8
1.8	137302	10242	10242	100.00%	114.20%	2.38	77.9
1.7	148074	10835	10869	99.70%	209.80%	1.22	47.8
Total	932033	67485	67530	99.90%	8.40%	19.77	100.0

Table 3.6 Scaling statistics for StiE 961-1257 (PDB ID 6ECT)

Resolution Limit (Å)	Number of Reflections						
	Observed	Unique	Possible	Completeness	R-Factor	I/σ	$CC_{1/2}$
4.23	49027	2773	2795	99.20%	3.60%	74.08	99.9
3	83208	4733	4746	99.70%	3.90%	69.92	99.9
2.45	114543	6023	6027	99.90%	4.70%	58.57	99.9
2.12	130611	7021	7080	99.20%	6.50%	45.2	99.9
1.9	149591	7823	7951	98.40%	8.90%	32.29	99.8
1.74	178094	8751	8757	99.90%	15.10%	18.85	99.6
1.61	189305	9489	9490	100.00%	25.50%	11.02	98.9
1.5	207264	10200	10200	100.00%	45.50%	6.51	96.8
1.42	175493	10303	10791	95.50%	82.20%	3.13	85.5
Total	1277136	67116	67837	98.90%	5.80%	27.23	100.0

Table 3.7 Scaling Statistics for StiE 942-1257 (PDB ID 6ECX)

Resolution Limit (Å)	Number of Reflections						
	Observed	Unique	Possible	Completeness	R-Factor	I/σ	$CC_{1/2}$
5.66	13816	1171	1175	99.70%	2.70%	78.8	100.0
4.02	24162	1951	1951	100.00%	3.00%	74.28	100.0
3.28	32049	2461	2461	100.00%	4.00%	56.49	99.9
2.85	37682	2880	2880	100.00%	7.50%	31	99.9
2.55	42484	3239	3239	100.00%	15.50%	16.05	99.5
2.33	46287	3546	3546	100.00%	29.40%	8.72	98.4
2.15	50993	3850	3850	100.00%	51.20%	5.1	95.1
2.01	53514	4122	4123	100.00%	105.30%	2.46	82.7
1.9	57371	4340	4357	99.60%	193.70%	1.28	59.0
Total	358358	27560	27582	99.90%	6.80%	21.18	100.0

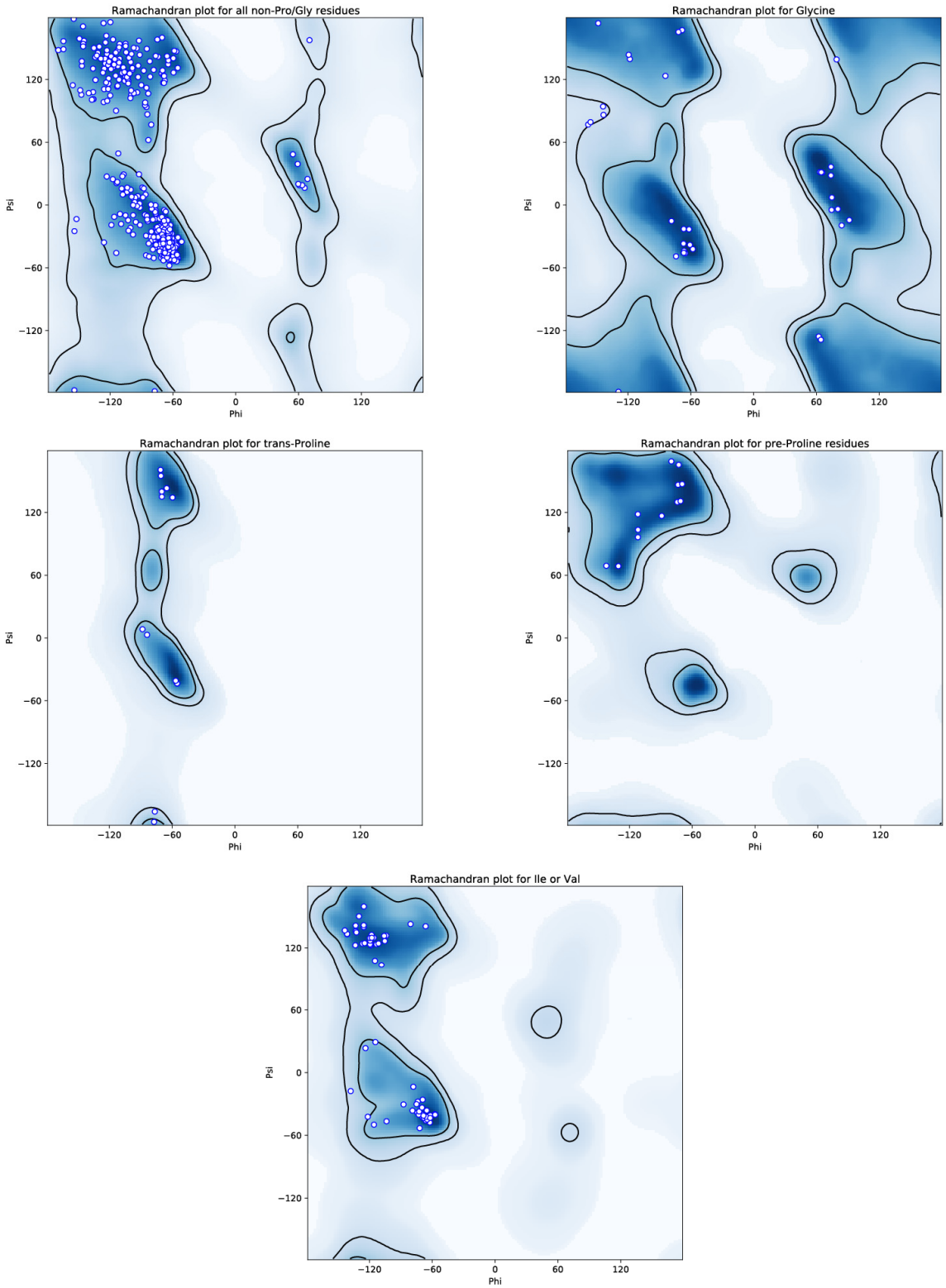


Figure 3.2 Ramachandran plots for SeMet labeled StiD *O*-MT 976-1266
Ramachandran plots of final refined model of SeMet labeled StiD *O*-MT 976-1266 (PDB ID 6ECU). Plots were generated using MolProbity[133].

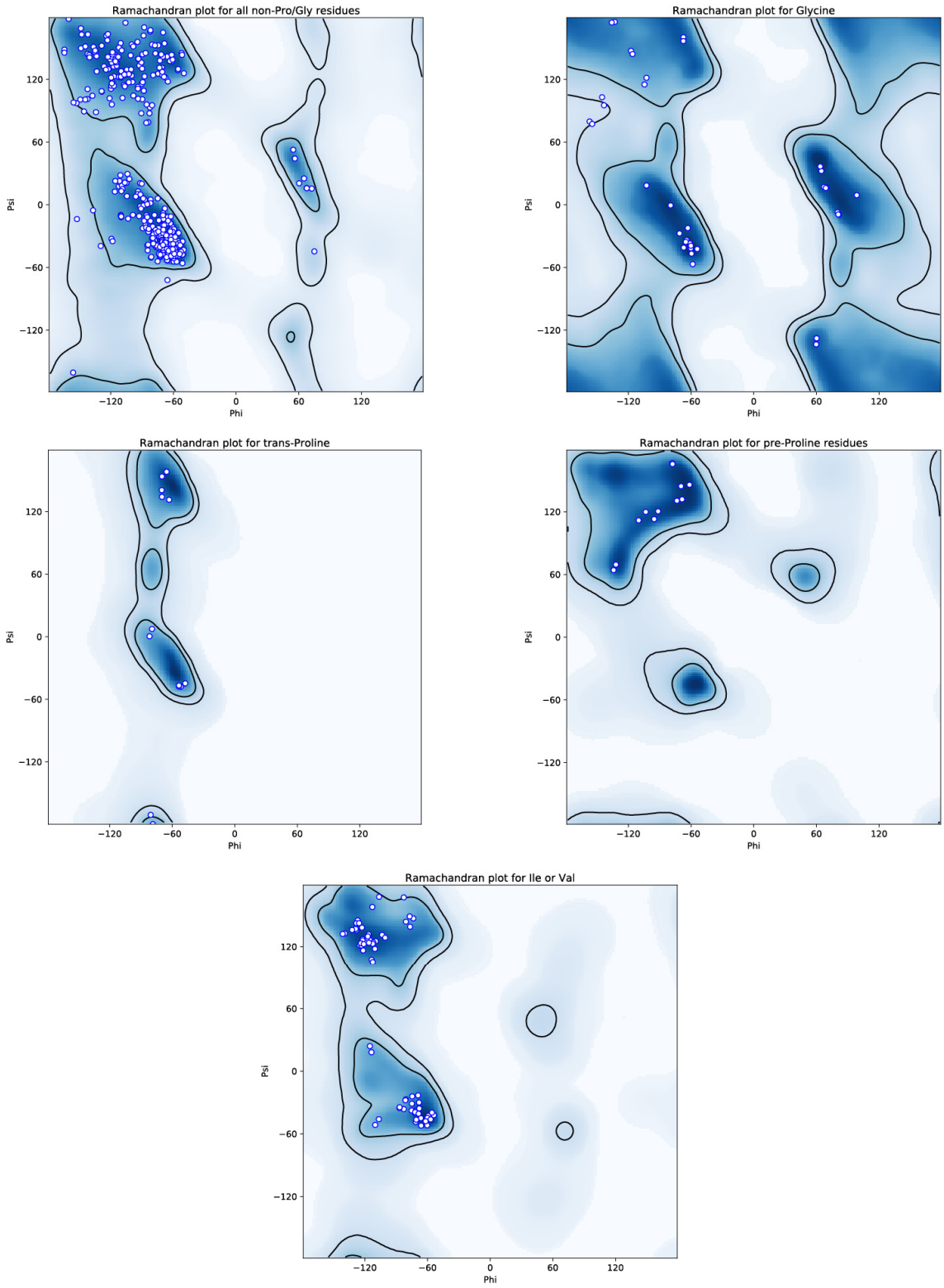


Figure 3.3 Ramachandran plots for native StiD *O*-MT 976-1266
Ramachandran plots of final refined model of native StiD *O*-MT 976-1266 (PDB ID 6ECV). Plots were generated using MolProbity[133].

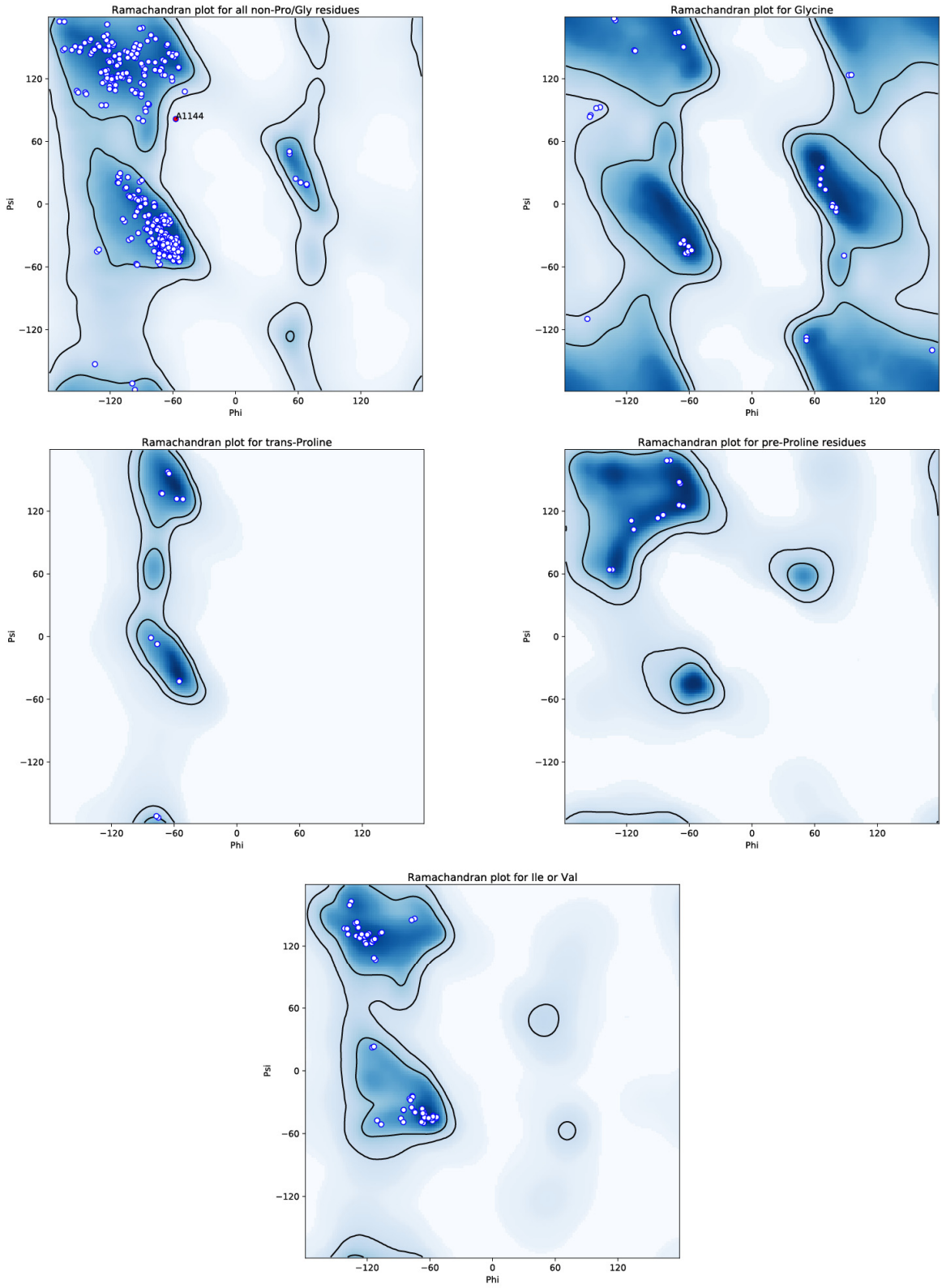


Figure 3.4 Ramachandran plots for StiD *O*-MT 956-1266
Ramachandran plots of final refined model of StiD *O*-MT 956-1266 (PDB ID 6ECW). Plots were generated using MolProbity[133].

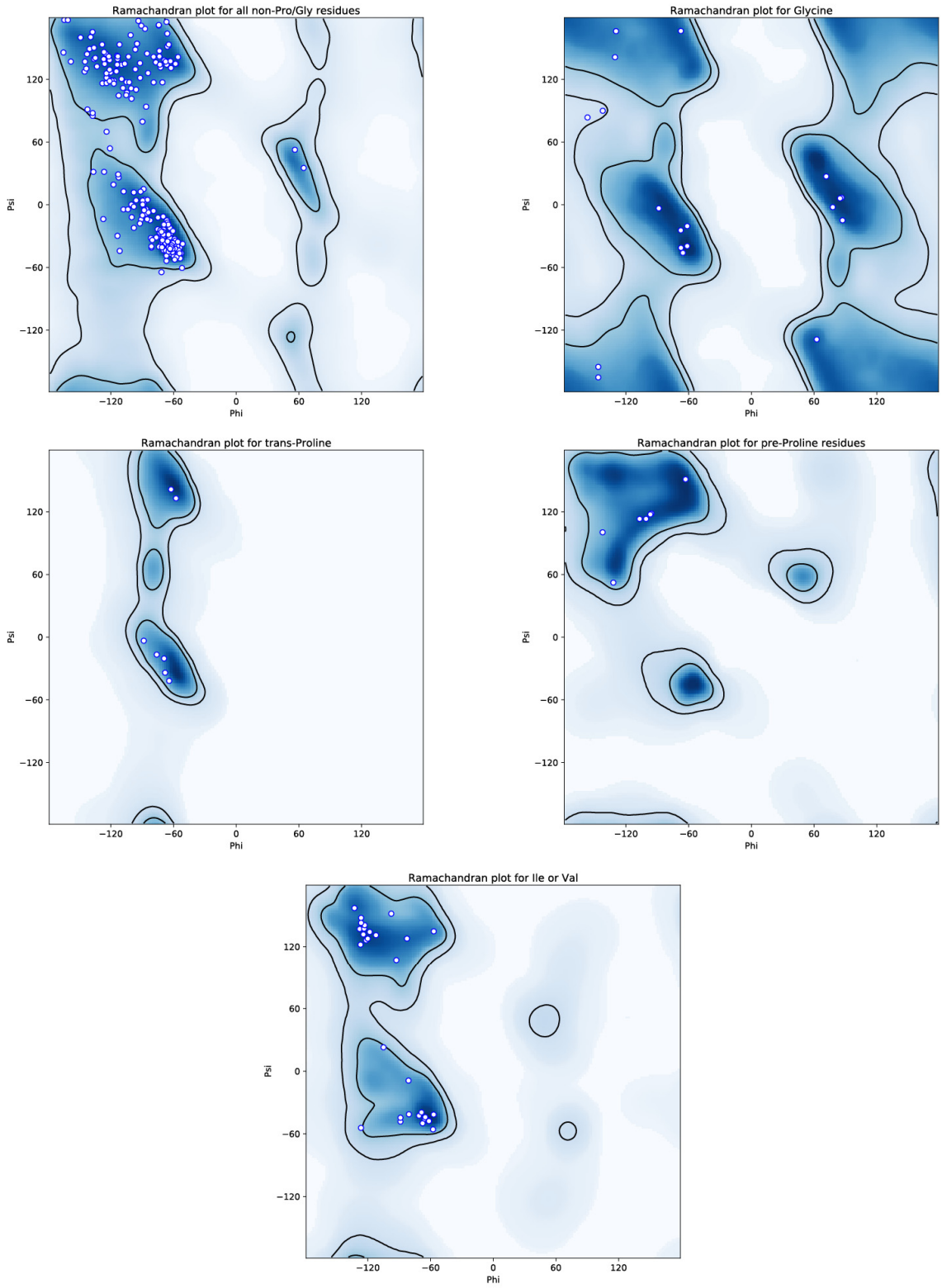


Figure 3.5 Ramachandran plots for StiE *O*-MT 961-1257
Ramachandran plots of final refined model of StiE *O*-MT 961-1257 (PDB ID 6ECT). Plots were generated using MolProbity[133].

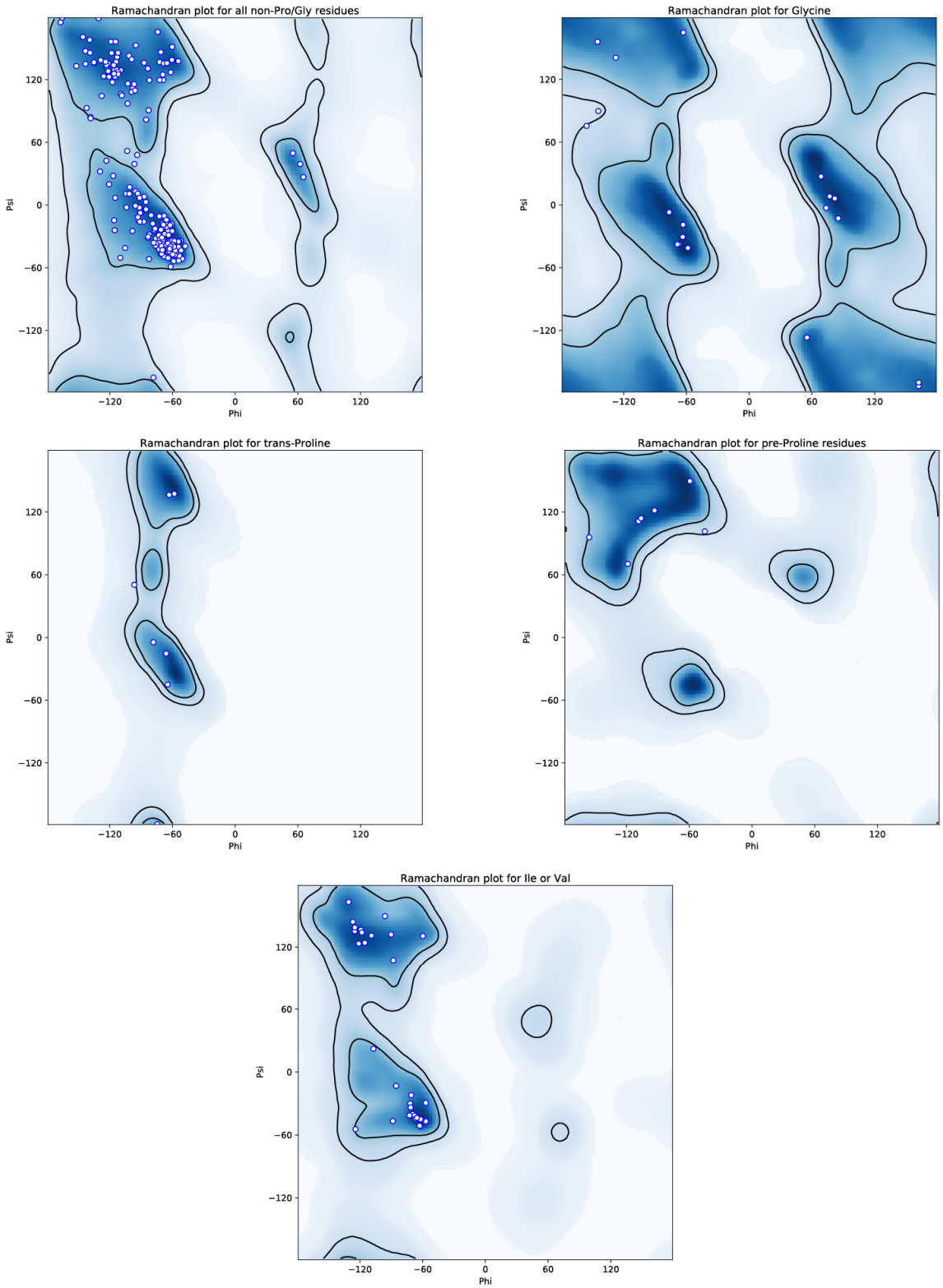


Figure 3.6 Ramachandran plots for StiE *O*-MT 942-1257
Ramachandran plots of final refined model of StiE *O*-MT 961-1257 PDB ID 6ECX). Plots were generated using MolProbity[133].

Results and Discussion

Determining O-MT domain boundaries

We first examined the *O*-MT domain boundaries to excise the domain from the module. *O*-MTs that methylate at a β-hydroxyl are flanked by AT and KR domains. Sequence motifs delineating the C-terminus of the AT (so-called post-AT linker) and N-terminal β-strand of the KR (Figure 3.7), guided the production of nine StiD *O*-MT constructs (Figure 3.8). Three StiD fragments were soluble and two (956-1266, 976-1266) were purified to homogeneity. Soluble fragments of StiE (942-1257, 961-1257) and CurL (981-1315) were designed based on the experimentally determined domain boundaries for StiD. StiD was predominantly dimeric in solution as determined by size exclusion chromatography, whereas StiE *O*-MT was predominantly monomeric and CurL *O*-MT was monomeric (Figure 3.8). Inclusion of an additional 30 amino acids between the CurL post-AT linker and *O*-MT (951-1315) (Figure 3.7) resulted in dimeric CurL *O*-MT (Figure 3.8), demonstrating that this is a post-AT dimerization element similar to those between the AT and KR in Spn module 3[52] and CurI.[48]

Table 3.8 Protein stability of StiD fragments containing the *O*-MT

	N-950	**N-956**	**N-976**
C-1245	insoluble	insoluble	insoluble
C-1257	insoluble	insoluble	insoluble
C-1266	soluble	soluble	soluble

O-MT activity and substrate specificity

We next tested the ability of excised StiD, StiE, and CurL *O*-MT to methylate mimics of their predicted substrates. When StiD *O*-MT was presented with a truncated triketide version of the presumed (*S*)-β-hydroxy substrate linked to *N*-acetylcysteamine (NAC), a common mimic for phosphopantetheine (Ppant), in the presence of *S*-adenosylmethionine (SAM) no methylation occurred. However, when presented with an excised StiD ACP-linked (*S*)-configured β-hydroxy triketide substrate (**1**), StiD *O*-MT produced the expected β-methoxy product (**2**) as detected by

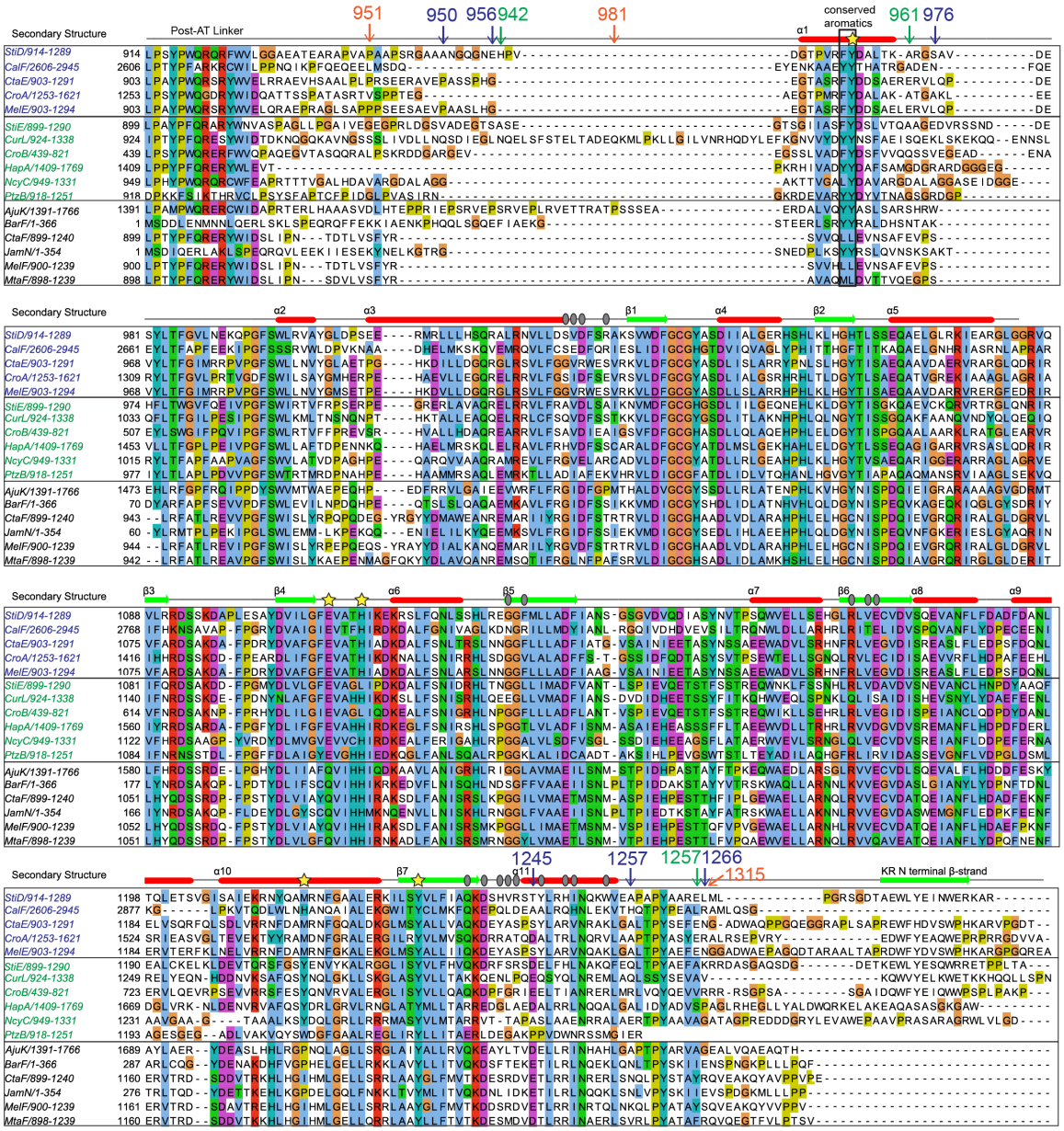


Figure 3.7 Sequence alignment of PKS *O*-MTs

Names are colored based on predicted substrate ((*S*)-OH, blue; (*R*)-OH, green; keto; black). Substrate predictions are based upon structures of the final metabolites, the inclusion of a KR in *O*-MT containing modules, and KR sequence motifs that correlate with the stereochemical configuration of the resulting methoxy. Stars indicate sites of mutagenesis; arrows represent experimental N- and C-termini tested for StiD (blue) and fragment N- and C-termini for StiE (green) and CurL (orange). Gray ovals represent residues in the StiD or StiE *O*-MT dimer interface. Pathway abbreviations (GenBank Accision codes) are as follows: Cal- calyculin A (BAP05594.1); Cta- cystothiazole A (AAW03329.1, AAW03329.1); Cro- crocacin (AIR74910, AIR74911.1); Mel- melithiazol (CAD89776.1, CAD89777.1); Sti- stigmatellin (CAD19088.1, CAD19089.1); Cur- curacin A (WP_008191786); Hap- haprolid (AOG74798.1); Nyc- nannocystin A (ALD82523.1); Ptz- patellazole (AFX99666.1); Aju- ajudazol (CAQ18838.1); Bar- barbamide (AAN32980.1); Jam- jamaicamide (AAS98785.1); Mta- myxothiazol (AAF19814.1).

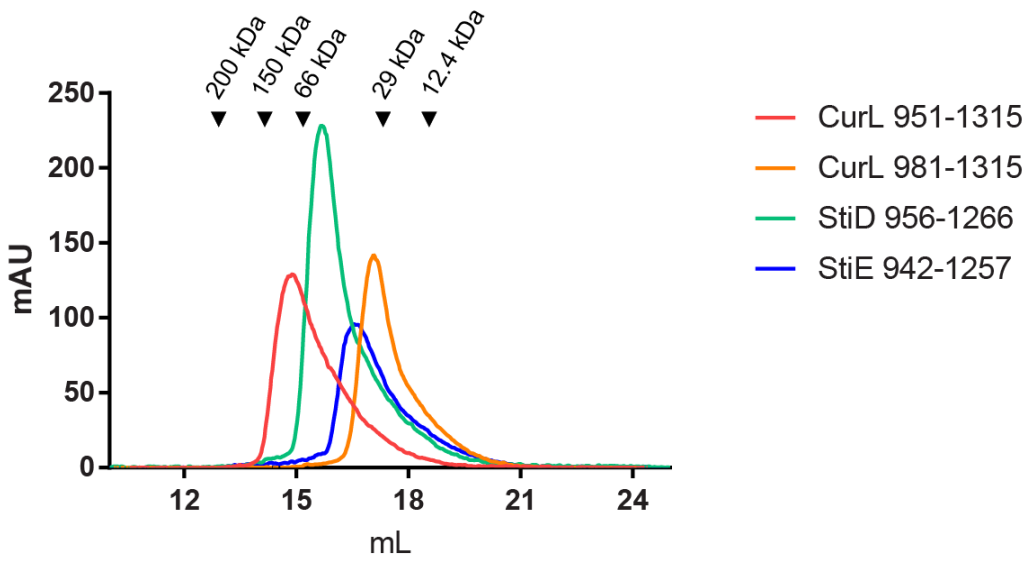


Figure 3.8 Oligomeric state of *O*-MTs
Size exclusion chromatography of *O*-MT fragments. StiD *O*-MT (amino acids 956-1266, 35 kDa monomer) elutes with an apparent molcular weight of 55 kDa and StiE *O*-MT (amino acids 942-1257, 37 kDa monomer) elutes predominantly as a monomer (apparent molecular weight 35.7 kDa). CurL *O*-MT (amino acids 981-1315, 39 kDa monomer) is exlusively monomeric (apparent molecular weight 26.9 kDa). Inclusion of a 30 residue post-AT dimerization element at the CurL N-terminus (amino acids 951-1315, 42 kDa monomer) results in dimeric protein (apparent molecular weight 82.8 kDa).

the mass spectrometry-based Ppant ejection assay (Figure 3.9c).[150, 169] Additionally StiE *O*-MT and CurL *O*-MT, which are predicted to act on (*R*)-β-hydroxy substrates (Figures 3.10a, c), methylated the respective (3*R*)-3-hydroxy-5-methoxy-myristoyl-ACP (**6**, Figure 3.10**)** and (*R*)-3-hydroxydodecanoyl-ACP (**8**, Figure 3.10) substrates (Figures 3.10b, d). Given the encoded stereospecificity of KR domains, we tested whether *O*-MTs select for stereocenters at the α- and β-positions. A series of ACP linked diastereomeric triketide substrates (**2**, **3**, **4,** Figure 3.9) with varying stereocenters at the α-methyl and β-hydroxyl groups were tested with StiD *O*-MT. StiD did not methylate any substrate with non-native stereocenters at the α- or β-positions (Figures 3.9d, 2e, 2f), indicating that PKS *O*-MTs display substrate stereospecificity.

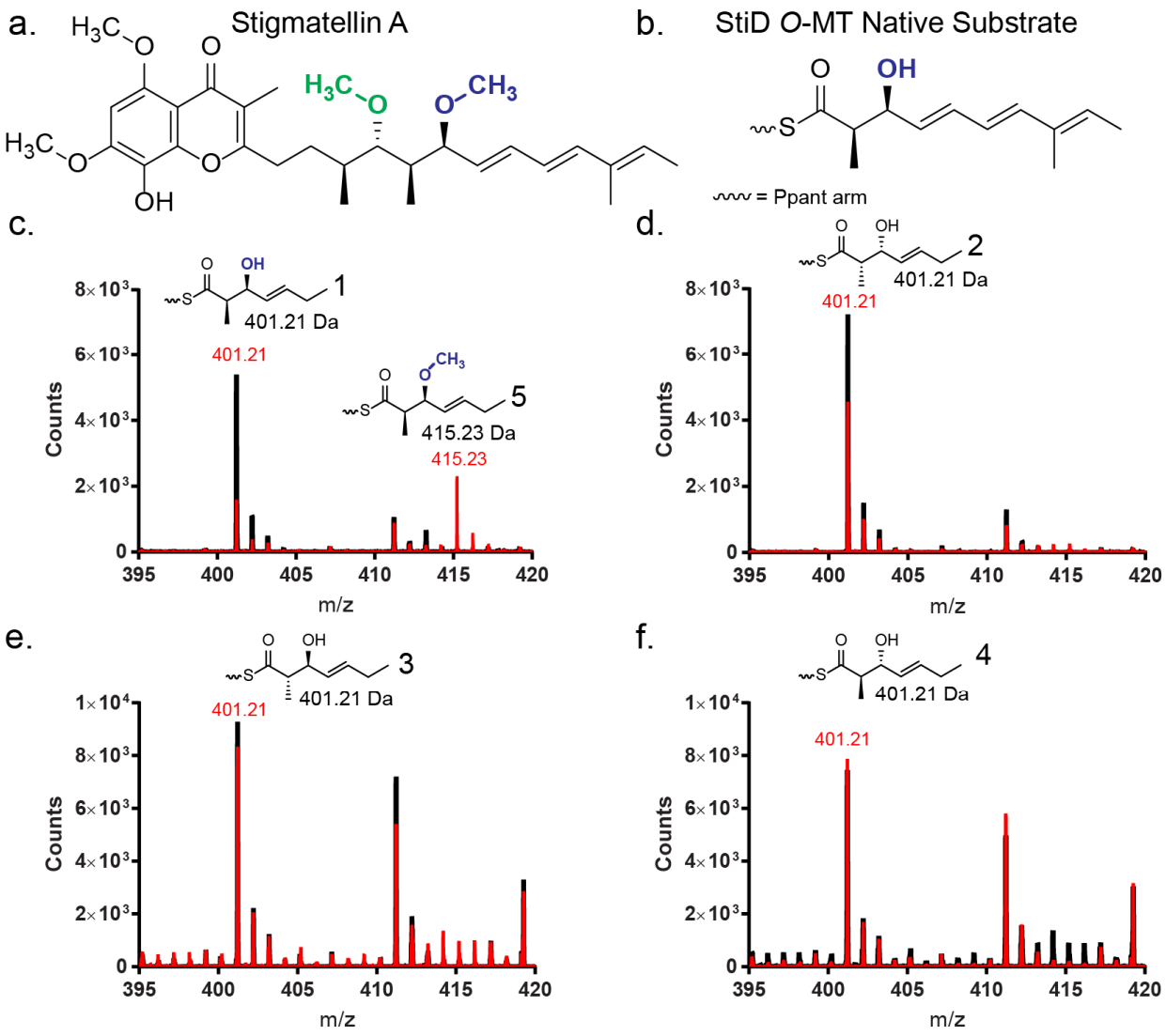


Figure 3.9 Activity of StiD *O*-MT on ACP linked substrates

a. The metabolite stigmatellin contains two methoxy groups with opposite stereochemical configurations. The methoxy moieties introduced by *O*-MTs encoded in PKS modules are colored (StiD, blue; StiE, green). **b.** Natural substrate of StiD *O*-MT. **c.-f.** Mass spectra of reaction mixtures of StiD *O*-MT with various triketide substrate mimics (**1**, **2**, **3**, **4**) linked to StiD ACP monitored by the phosphopantetheine ejection assay. [150, 169] Reactions are shown in red, no enzyme controls are shown in black. Experimentally determined m/z values are in red, calculated m/z values are in black. A contaminant of m/z ~411 Da was present in all samples.

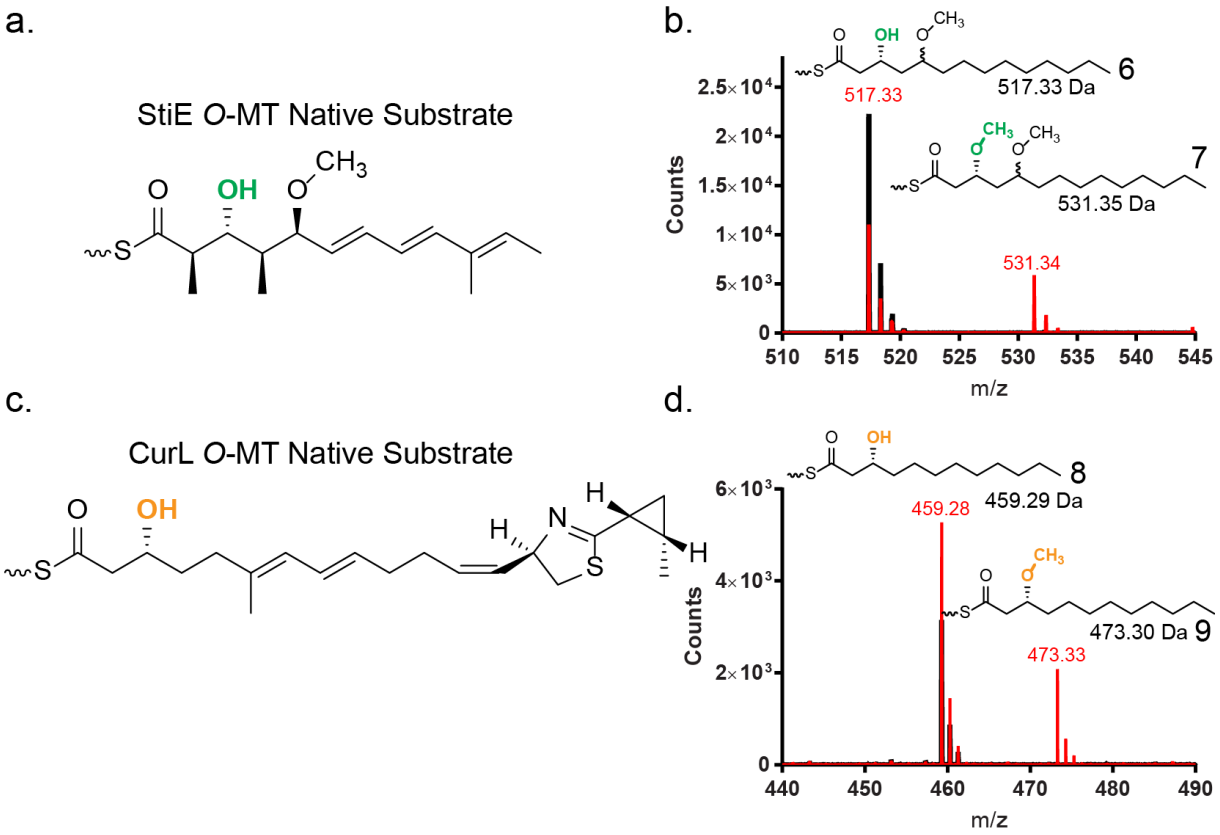


Figure 3.10 Activity of StiE and CurL *O*-MTs on ACP linked substrates
a. Natural substrate of StiE *O*-MT. **b.** Mass spectra of reaction mixtures with StiE *O*-MT and StiE (3*R*)-3-hydroxy-5-methoxy-myristoyl-ACP substrate mimic (**6**). **c.** Natural substrate of CurL *O*-MT. **d.** Mass spectra of reaction mixtures of reactions with CurL *O*-MT and CurL (*R*)-3-hydroxydodecanoyl-ACP substrate mimic (**8**). Reactions are shown in red, no enzyme controls are shown in black. Experimentally determined m/z values are in red, calculated m/z values are in black. A contaminant of m/z ~411 Da was present in all samples.

Structural characterization of StiD and StiE O-MT

To understand the structural basis of *O*-MT reactivity, we determined crystal structures of StiD and StiE *O*-MT with varying domain boundaries (Figure 3.11, Table 3.2). An initial structure of selenomethionine-labeled StiD *O*-MT (971-1266) was solved by single-wavelength anomalous diffraction (SAD) phasing, and this structure was used to solve subsequent StiD and StiE *O*-MT structures by molecular replacement. StiD and StiE *O*-MTs (Figure 3), which methylate opposing stereocenters, are highly similar in structure (RMSD 0.88 Å for 193 Cα atoms). PKS *O*-MTs are members of the class I methyltransferase superfamily, which share a seven β-stranded SAM binding core.[125] The StiD dimer interface is mediated by hydrophobic and hydrogen bonding contacts (StiD Phe1134, Arg1173, Val1175, Glu1176, Gln1237, Asp1239,

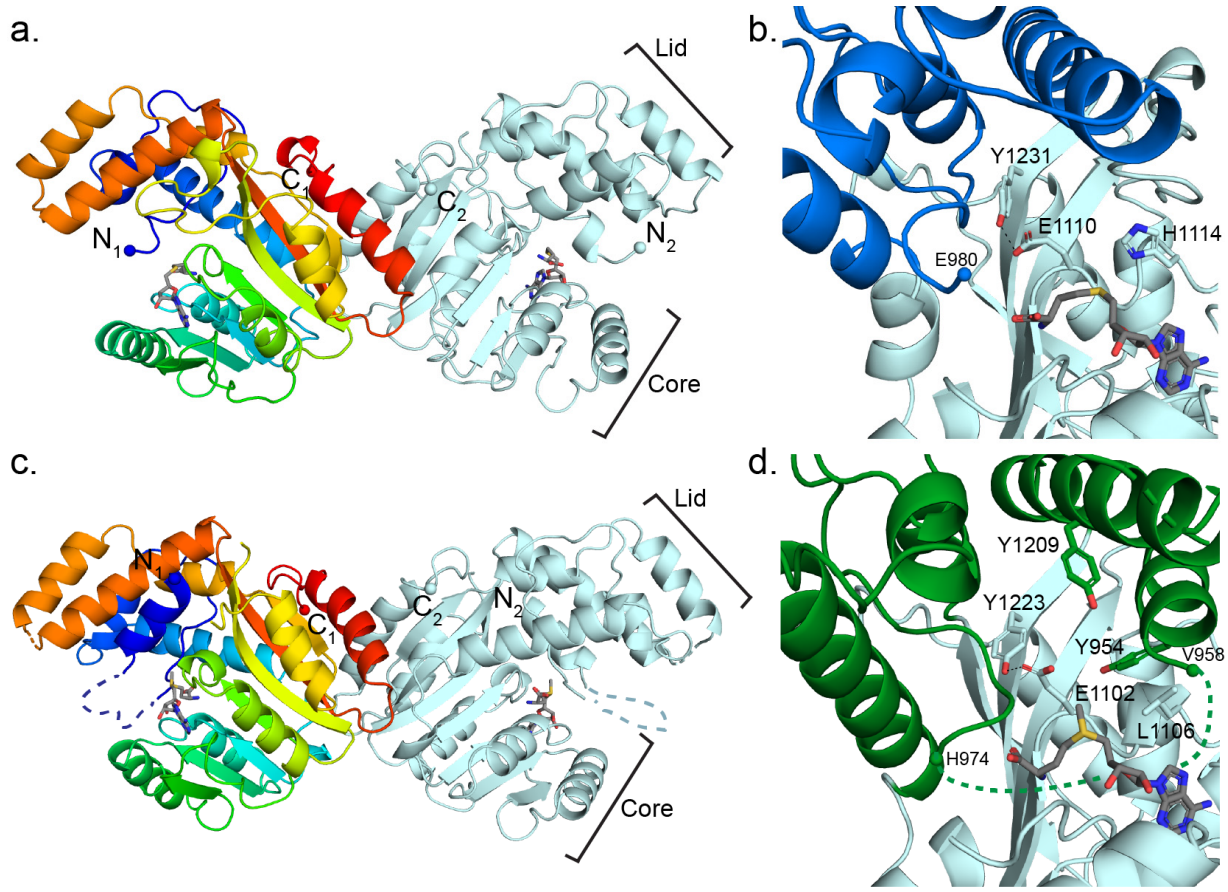


Figure 3.11 StiD and StiE *O*-MT structures.
a. StiD *O*-MT dimer. One protomer is colored as a rainbow from blue (N-terminus) to red (C-terminus). SAH is shown in sticks with atomic coloring (C, gray; O, red; N, blue; S, yellow). Termini are shown in spheres. **b.** StiD *O*-MT active site colored by structural region (core, light blue; lid, dark blue). Key amino acids and SAH are shown in sticks with atomic coloring. The first ordered residue in the crystal structure (Glu980) is indicated with a sphere. **c.** StiE *O*-MT dimer colored as in a. SAM is shown in sticks. **d.** StiE *O*-MT active site colored by structural region (core, light blue; lid, green). Key amino acids and SAM are shown in sticks with atomic coloring. The last amino acid (Val958) of the N-terminal helix containing conserved aromatic amino acid Tyr954 is 17 Å away from the next ordered residue (His974).

Val1242, Tyr1246, His1249, Ile1250, Trp1254). Although predominantly monomeric in solution, StiE forms a dimer in the crystal with the analogous interaction interface as StiD (StiE Ala1021, Asp1023, Ala1026, Gly1124, Leu1126, Arg1165, Val1167, Gln1229, Asp1231, Phe1233, Arg1235) (Figure 3.12), suggesting that StiE is dimeric in the context of a module. The angle between monomers within the StiD and StiE dimers differs by ~20°. CurL *O*-MT, which did not crystallize, has several amino acid differences at surface analogous to the StiD and StiE dimer

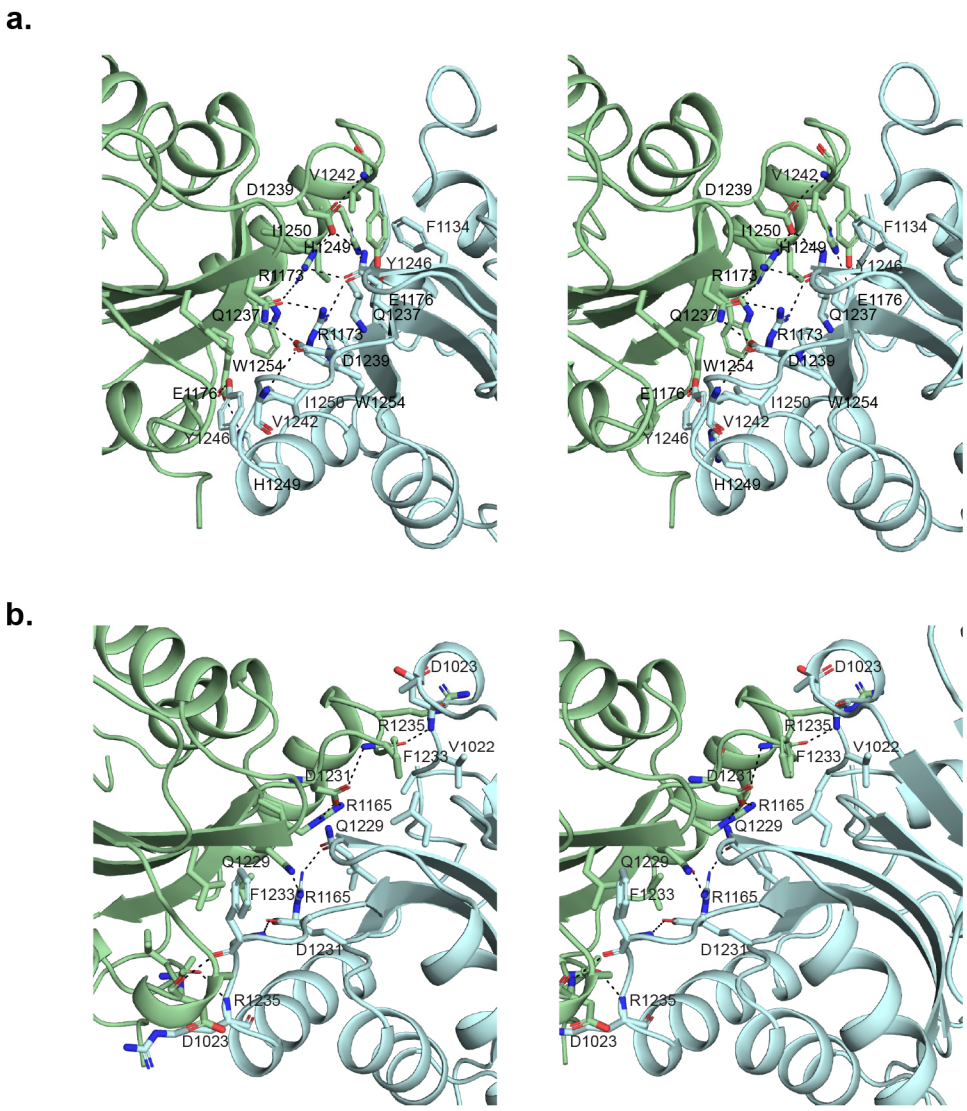


Figure 3.12 StiD and StiE *O*-MT dimer interfaces
a. The StiD *O*-MT dimer interface is mediated by amino acids Phe1134, Arg1173, Val1175, Glu1176, Gln1237, Asp1239, Val1242, Tyr1246, His1249, Ile1250, Trp1254 from each monomer. **b.** StiE *O*-MT dimer interface is mediated by amino acids Ala1021, Asp1023, Ala1026, Gly1124, Leu1126, Arg1165, Val1167, Gln1229, Asp1231, Phe1233, Arg1235.

interface indicating that CurL *O*-MT may rely on the dimerization element for dimerization within the module (Figure 3.7).

Different branches of the class I MT superfamily can be distinguished by a lid domain, which lies over the common SAM binding site and is associated with substrate binding.[65] StiD and StiE *O*-MT have a helical lid composed of two N-terminal helices and three helices inserted between β-strands six and seven of the MT core (Figure 3.7). The lid has substantially different positions (up to 8 Å displacement) in the six independent views of the StiD protomer within three crystal structures (Figure 3.13), indicating that it is dynamic in the substrate-free state.

Despite a low sequence identity (<20%), the PKS *O*-MTs have greatest structural similarity to the phosphoethanolamine *N*-methyltransferase (PDB 3UJ8, RMSD 1.9 Å for 153 Cα atoms),[175] mycolic acid cyclopropane synthase (PDB 1KPG, RMSD 1.6 Å for 152 Cα atoms),[149] RebM sugar *O*-methyltransferase (PDB 3BUS, RMSD 1.6 Å for 145 Cα atoms),[68] and the geranyl diphosphate *C*-methyltransferase (GPPMT) (PDB 3VC2, RMSD 1.4 Å for 155 Cα atoms).[72] Class I MTs on this branch of the superfamily have helical lids of similar structure, including helices at the N-terminus and inserted between β-strands 6 and 7 of the core (Figure 3.14). Interestingly, this MT superfamily branch includes the SpnF Diels-Alderase (PDB 4PNE, RMSD 3.5 Å for 193 Cα atoms), which was adapted to a new function and does not catalyze methyl transfer[176].

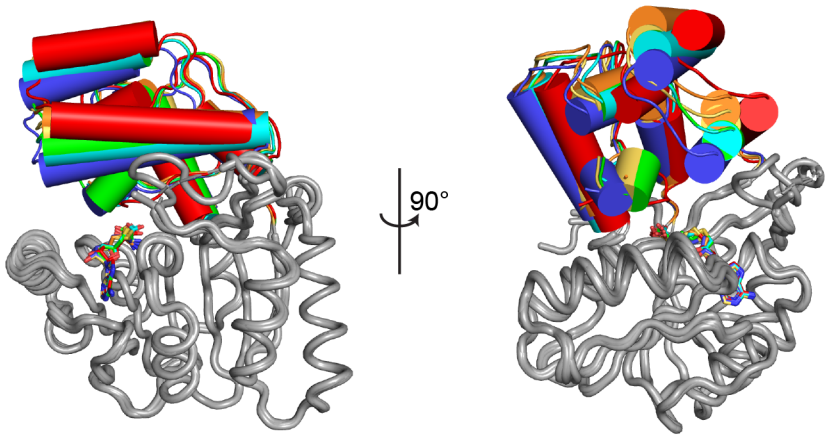


Figure 3.13 StiD lid dynamics
StiD structures aligned at core (123 Cα atoms, gray tube). Regions of the lid differ in position by up to 8 Å.

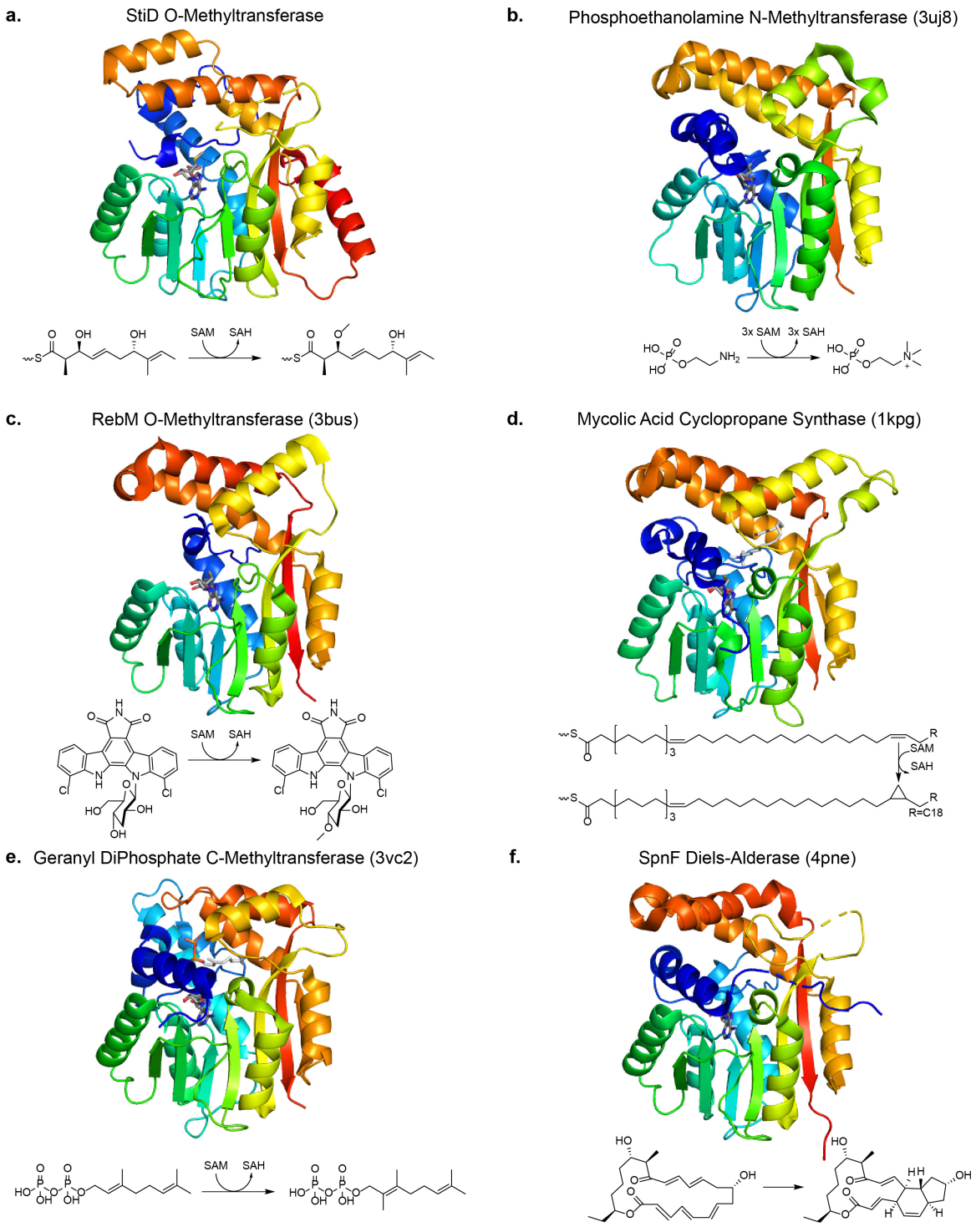


Figure 3.14 PKS *O*-MT homologs

Structures are colored as a rainbow from N- (dark blue) to C-terminus (red). The reaction catalyzed by each enzyme is shown below the structure. All structures have a lid composed of N-terminal helices (darkest blue) and helices between β-strands 6 and 7 (yellow and orange). SAH or SAM is rendered in stick form with gray C atoms.

Probing O-MT active site architecture

The StiD 976-1266 and StiE 961-1257 *O*-MT structures have solvent-exposed active sites and also lack an N-terminal lid helix that lays over the SAM binding site in the structures of several homologs (Figure 3.14). The analogous site above SAM in the StiE 961-1257 structure is occupied by a cloning-artifact peptide from the TEV protease recognition site (Tyr-Phe-Gln-Ser-Asn-Ala, Figure 3.15a). The peptide sequence is similar to a sequence in StiE (^{953}Phe-Tyr-Asp-Ser-Leu-Ala^{958}) that precedes the crystallized fragment (961-1257). Moreover, the PKS *O*-MT sequences contain adjacent conserved hydrophobic (preferred aromatic) amino acids in the N-terminal region (Figure 3.7). Thus, we incorporated additional N-terminal amino acids in the StiD and StiE fragments and removed the TEV protease recognition sequence for additional crystallographic studies, resulting in a 1.7-Å structure of StiD 956-1266 and a 1.9-Å structure of StiE 942-1257. No density was observed for the additional N-terminal residues of StiD. However, the map for StiE 942-1257 had density for a partially ordered helix above the SAM binding site (Figure 3.15b). We modeled residues 948-958, including conserved Phe953-Tyr954, into the helical density. The helix is 17 Å away from the next ordered amino acid, His974 (Figure 3.11d).

We used site directed mutagenesis to test whether the conserved N-terminal aromatic amino acids play a role in catalytic activity of the robustly active StiE and CurL *O*-MTs, which

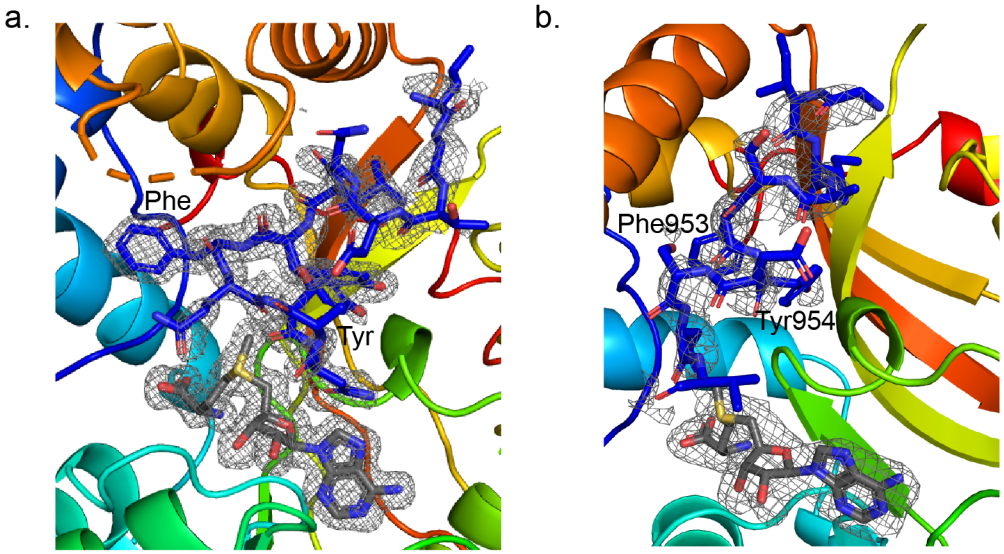


Figure 3.15 StiE active site density
a. StiE *Fo*-*Fc* omit density for the SAM cofactor and TEV protease recognition sequence (StiE 961-1257, 1.42 Å, 2.5σ contour). Phe and Tyr in the TEV protease recognition sequence are labeled. **b.** SAM and the partially ordered N-terminal helix (StiE 942-1257, 1.90 Å, *Fo*-*Fc* omit density contoured at 2.5σ). Conserved Phe953 and Tyr954 in the N-terminal helix are labeled.

methylate (*R*)-β-hydroxy substrates (Figures 3.16b, Figure 4c). Relative activity was quantified based upon the abundance of Ppant ejection fragments of the substrates and products[150, 169]. A CurL Y1010F variant had barely detectable activity (50-fold less than wild-type) (Figure 3.16c). In contrast, a StiE Y954F variant (Figure 3.16a) had increased activity relative to the wild type, but deletion of the StiE N-terminal helix (StiE Δ942-960) abolished detectable activity (Figure 3.16b). Although the Tyr to Phe substitutions had differing effects on the catalytic activity of StiE and CurL *O*-MTs, we conclude that the presence of the N-terminal helix is critical for activity and the conserved aromatic residues are likely near or in the active site when substrate is bound. Additional evidence comes from the PKS *O*-MT homolog phosphoethanolamine *N*-MT, which contains an essential Tyr in the analogous N-ter minal helix.[175] We hypothesize that the N-terminal helix containing the conserved aromatic region may order upon delivery of the ACP-bound substrate and serve as a "latch" to close the active site and facilitate catalysis.

SAM or the product (*S*)-adenosylhomocysteine (SAH) bound to both StiD and StiE in a similar orientation as in the structures of homologs (Figures 3.11, Figure 3.17, Table 3.2). In some of these structures, a His side chain adjacent to the SAM methyl is associated with activity (Figure 3.11b, Figure 3.7, Figure 3.17).[68, 175] The His is not conserved in PKS *O*-MTs, as the corresponding position in StiE is Leu1106 (Figure 3.11, Figure 3.16). The CurL variants H1165A and H1165N had 10- to 20-fold decreased activity relative to wild type (Figure 3.16c), and StiE L1105H displayed a two-fold loss in activity (Figure 3.16b), indicating that the active site His may assist in substrate binding, but is not essential for catalysis.

Two amino acids in the active site are conserved among PKS *O*-MTs. In StiE, conserved Glu1102 (StiD Glu1110, CurL Glu1161) is positioned with one carboxylate oxygen 4 Å from the SAM methyl group and the other carboxylate oxygen hydrogen bonded with the hydroxyl of invariant Tyr1223 (StiD Tyr1231; CurL Tyr 1281) (Figure 3.7, Figures 3.11b, d, Figure 3.16a). Several PKS *O*-MT homologs have a Glu or Asp within hydrogen bonding distance of a Tyr or Trp at the analogous positions (Figure 3.17). Substitutions of the Glu or Tyr were highly deleterious to catalysis in both StiE and CurL *O*-MTs (StiE E1102A, E1102Q,

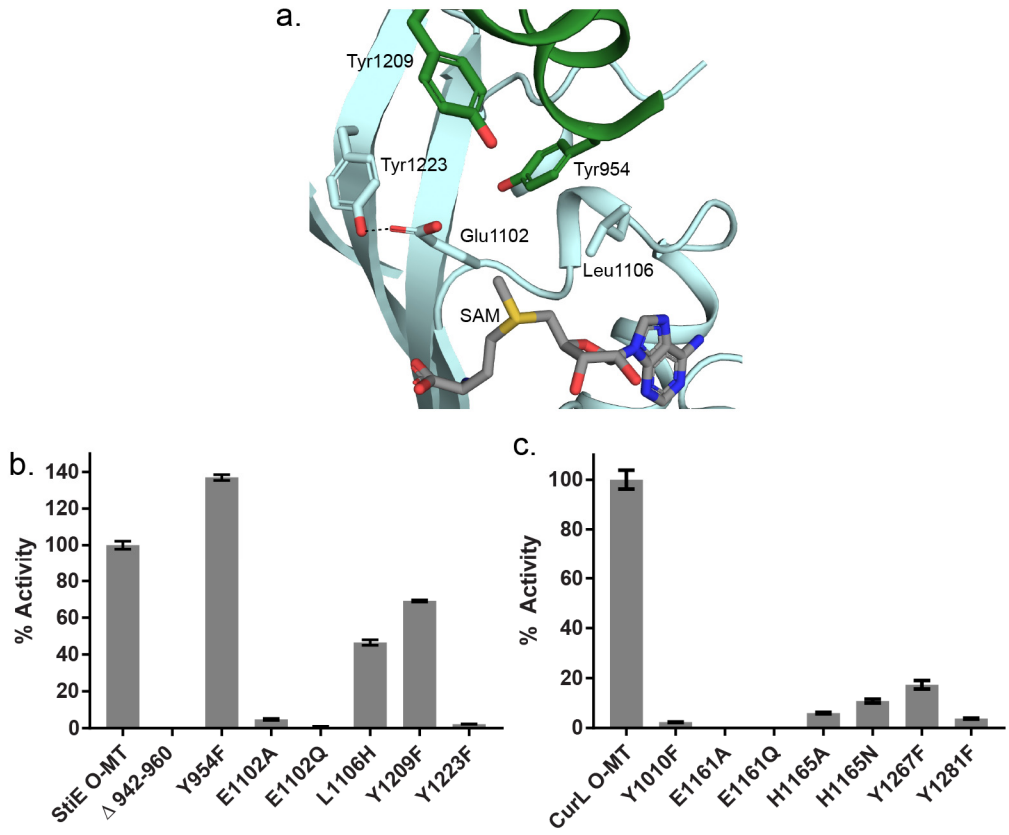


Figure 3.16 Relative methylation activities of CurL and StiE *O*-MT variants
a. StiE *O*-MT active site colored as in Fig. 3d. Substituted amino acids are shown in sticks. **b.** StiE *O*-MT activity on (3*R*)-3-hydroxy-5-methoxy-myristoyl-ACP substrate (**6**, Figure 3.10). **c.** CurL *O*-MT activity on (*R*)-3-hydroxydodecanoyl-ACP substrate (**8**, Figure 3.10). Activity relative to wild-type was quantified using the LC-MS based Ppant ejection assay.[150, 169] The mutagenesis sites in StiE and CurL *O*-MTs are analogous (StiE/CurL: Tyr954/Tyr1010, Glu1102/Glu1161, Leu1106/His1165, Tyr1209/Tyr1267, Tyr1223/Tyr1281) based on alignment of the 35% identical sequences. Error bars represent triplicate experiments
.

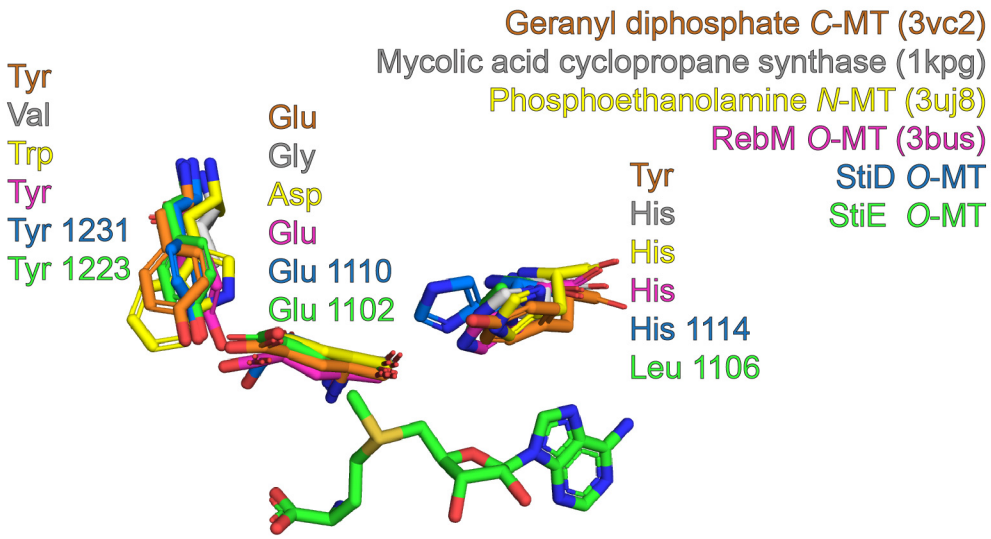


Figure 3.17 Conservation of key active site residues in *O*-MT homologs
Many *O*-MT homologs have similar active site features. Several homologs contain a Glu or Asp in an analogous location to PKS *O*-MTs catalytic Glu. Like PKS *O*-MTs, the Glu is hydrogen bonded to a Tyr.

Y1223F; CurL E1161A, E1161Q, Y1281F) (Figures 3.16b, c), suggesting that the conserved Glu-Tyr pair is essential for catalysis in all PKS *O*-MTs that methylate β-hydroxyl groups. The proximity to the SAM methyl makes Glu the primary candidate for deprotonation of the substrate β-hydroxyl during the methyl transfer reaction.

Finally, a Tyr just outside the active site of StiE *O*-MT (Tyr1209, Figure 3.16a) occurs in many *O*-MTs that are predicted to methylate *R*-hydroxyl substituents, but not those acting on *S*-hydroxyl or β-keto groups (Figure 3.7). Substitution of the Tyr (CurL Y1267F; StiE Y1209F) resulted in a six-fold decrease in CurL activity, but only had a modest effect on StiE activity. Although conserved in most *R*-hydroxyl methylating *O*-MTs, this residue is not essential for StiE activity.

Catalytic strategy of PKS O-MT domains

Based on the mutagenesis results and sequence comparison, we propose that binding of an ACP-linked substrate triggers ordering of the N-terminal helix to form the Michaelis complex by positioning the appropriate substrate oxygen atom adjacent to the SAM methyl group and active site Glu. The conserved Glu-Tyr pair (Figures 3.11b, d, Figure 3.16) is essential for activity and likely deprotonates the β-hydroxyl. Interestingly, *O*-MTs predicted to act on β-keto substrates (Figure 3.1c), and thus not requiring β-hydroxy deprotonation, have a Gln in the position analogous

to the essential Glu (Figure 3.7). We tested whether a Gln substitution for the active site Glu would convert the β-hydroxyl methylating StiE *O*-MT or CurL *O*-MT to a β-keto methylase. However, neither StiE E1102Q nor CurL E1161Q methylated the simple β-keto substrate acetoacetyl-ACP. Additional active site amino acids that we found to have non-essential roles in StiE or CurL *O*-MT activity may play a role in substrate recognition and stereospecificity.

O-MTs in the context of a PKS module

The structural differences between PKS *O*-MTs and *C*-MTs are quite striking. Although both contain the seven β-stranded SAM binding core, their lid domains are topologically different and are inserted into different regions of the seven β-stranded MT core. Additionally, insertion of *O*-MTs and *C*-MTs into PKS modules differs. Like the "pseudo" methyltransferase in the PKS homolog metazoan fatty acid synthase (mFAS),[1] PKS *C*-MTs are embedded after the first N-terminal β-strand of the KR domain[117], whereas the *O*-MT precedes all elements of the KR. The *O*-MT seems to be a later addition in the evolution of PKS modules, whereas the last PKS-mFAS common ancestor likely contained a functional *C*-MT. As PKS modules are obligate dimers, the *O*-MT may assist dimer formation in PKS modules lacking DH domains. This is evident in the existence of small dimerization element that precedes the monomeric CurL *O*-MT (981-1315) (Figure 3.8). The presence of a dimerization element prior to the *O*-MT in cyanobacterial CurL *O*-MT, but not the myxobacterial *O*-MTs from the stigmatellin pathway and suggest that the overall architecture of PKS modules with KS-AT-*O*MT-KR-ACP could differ between pathways.

O-MTs are essential biosynthetic tools for the production of *O*-methoxy groups in polyketides. Identification of domain boundaries and amino acids necessary for catalysis provides fundamental information for the inclusion of *O*-MTs in PKS engineering efforts. *O*-MTs could be particularly valuable biocatalysts, as they are capable of selectively modifying hydroxyl groups prior to downstream synthetic transformations. Additionally, this information allows for the ablation of pathway *O*-MT activity by mutagenesis, providing a potential route to *O*-desmethyl polyketide products while maintaining the structural integrity of the PKS module.

Chapter 4 A Mononuclear Iron-Dependent Methyltransferase Catalyzes Initial Steps in Assembly of the Apratoxin A Polyketide Starter Unit

Reproduced with permission from Skiba, M.A., Sikkema, A.P., Moss, N.A., Tran, C.L., Sturgis, R.M., Gerwick, L., Gerwick, W.H., Sherman, D.H., Smith, J.L. A Mononuclear Iron-Dependent Methyltransferase Catalyzes Initial Steps in Assembly of the Apratoxin A Polyketide Starter Unit. *ACS Chem Biol* **12**, 3039-3048 (2017).

Summary

Natural product biosynthetic pathways contain a plethora of enzymatic tools to carry out difficult biosynthetic transformations. Here we discover an unusual mononuclear iron-dependent methyltransferase that acts in the initiation steps of apratoxin A biosynthesis (AprA MT_L). Fe^{3+}-replete AprA MT_L catalyzes one or two methyl transfer reactions on the substrate malonyl-ACP (acyl carrier protein), whereas Co^{2+}, Fe^{2+}, Mn^{2+} and Ni^{2+} support only a single methyl transfer. MT_L homologs exist within the "GNAT" (GCN5-related *N*-acetyltransferase) initiation modules of several modular biosynthetic pathways with propionyl, isobutyryl or pivaloyl starter units. GNAT domains are thought to catalyze decarboxylation of Mal-CoA and acetyl transfer to a carrier protein. In AprA the GNAT domain lacks both decarboxylation and acyl transfer activity. A crystal structure of the AprA MT_L -GNAT didomain with bound Mn^{2+}, malonate and the methyl donor *S*-adenosylmethionine (SAM) reveals that the malonyl substrate is a bidentate metal ligand, indicating that the metal acts as a Lewis acid to promote methylation of the malonyl α-carbon. The GNAT domain is truncated relative to functional homologs. These results afford an expanded understanding of MT_L -GNAT structure and activity, and permit the functional annotation of homologous GNAT initiation modules both with and without methyltransferases, additionally revealing their rapid evolutionary adaptation in different biosynthetic contexts.

Introduction

Prokaryotic secondary metabolite biosynthetic pathways use an assortment of enzymatic tools to synthesize some of the most elegant and chemically complex molecules found in nature.

Polyketide metabolites are mostly synthesized through the successive condensation of malonyl (Mal-) or methylmalonyl (MeMal-) coenzyme A (CoA) extender units in a manner similar to fatty acid biosynthesis. In type I modular polyketide synthase (PKS) systems, each extension step starts with acyltransferase (AT) selection of a specific carboxyacyl-CoA for transfer to the phosphopantetheine (Ppant) cofactor of the acyl carrier protein (ACP) domain within the module. This is followed by decarboxylative condensation with the upstream intermediate in the ketosynthase (KS) domain, and then by optional modifications by ketoreductase, dehydratase, enoylreductase or methyltransferase (MT) domains in the module, or external enzymes encoded in the gene cluster [28]. In this manner PKS systems synthesize a variety of essential pharmaceuticals, such as the anthelmintic avermectin, immunosuppressant rapamycin, and antibiotic erythromycin, using a common set of enzymes [106, 151, 177]. In some cases, PKS modules are interspersed with non-ribosomal peptide synthetase (NRPS) modules, which incorporate amino acid building blocks into the linear product.

Modular PKS systems initiate biosynthesis in a "initiation" or "loading" module by transfer of a specific acyl group to the loading ACP. Several mechanisms of pathway initiation have been described; frequently, a loading acyltransferase (AT_L) primes the ACP with an acyl group such as acetyl-, Mal-, or MeMal-CoA. An α-carboxylated acyl group can be decarboxylated by a non-extending ketosynthase domain [41]. In other cases, the initiation modules select chemically diverse starter units derived from amino acids, cyclohexanecarboxylic acids, fatty acids, or branched chain CoAs [41, 178]. For example, isobutyryl-CoA, methylbutyryl-CoA, and isovaleryl-CoA, derived from the catabolism of valine, leucine, and isoleucine through a branched chain α-keto acid dehydrogenase (BCDH) complex, are the starter units in the biosynthesis of virginiamycin, avermectin and myxothiazol, respectively [113, 179-181]. An AT_L directly primes the ACP with the branched-chain acyl group in these pathways [107, 151, 182].

We previously reported the discovery and characterization of the GNAT initiation module family (Figure 4.1). The CurA initiation module [31] initiates biosynthesis of the anti-cancer cyanobacterial compound curacin A [23]. Although the GNAT superfamily generally includes CoA-dependent *N*-acetyltransferases [42, 43], the CurA GNAT is a remarkable bifunctional enzyme that decarboxylates Mal-CoA to acetyl-CoA and catalyzes *S*-acetyltransfer to an associated ACP [31]. Clues to the evolutionary origin of the CurA GNAT decarboxylase activity came from the subsequent discovery that Mal-CoA decarboxylase (MCD) is a GNAT superfamily member [183].

MCD, which is conserved from bacteria to humans, catalyzes only decarboxylation. Thus, the GNAT superfamily may be better described as a "CoA binding" family, as members have a common mode of CoA binding and may catalyze acyltransfer, decarboxylation or both.

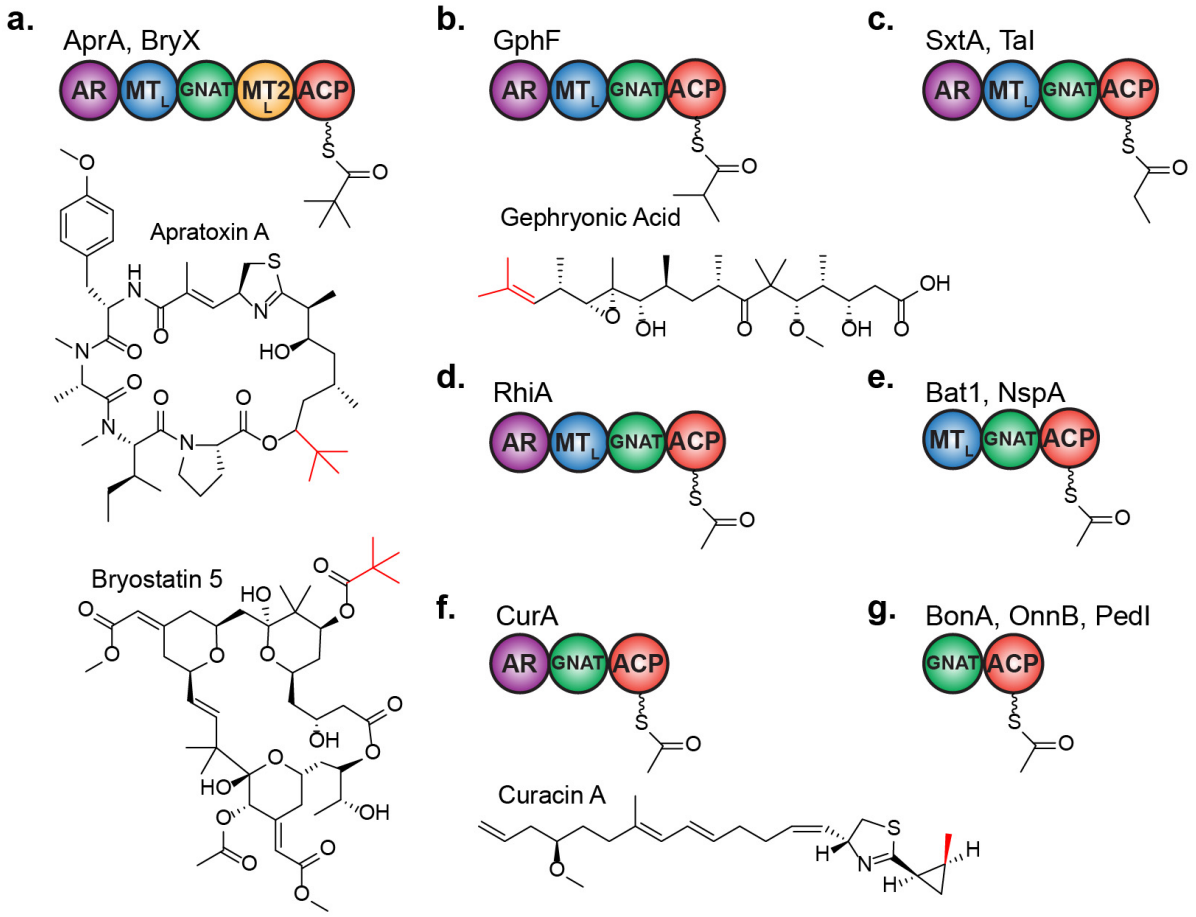


Figure 4.1 Domain architecture of GNAT containing initiation modules
Adaptor region (AR), methyltransferase (MT), GCN5-related *N*-acetyltransferase (GNAT), acyl carrier protein (ACP). Starter units produced by GNAT initiation modules in select metabolites are colored in red. Pathway abbreviations are as follows: **a.** Apr- apratoxin A, Bry- bryostatin. **b.** Gph- gephyronic acid. **c.** Sxt- saxitoxin, Ta- myxovirescin A. **d.** Rhi- rhizoxin. **e.** Bat- batumin, Nsp- nosperin. **f.** Cur- curacin A. **g.** Bon- bongkrekic acid, Onn- onnamide, Ped- pederin.

GNAT initiation modules in biosynthetic pathways for natural products that contain branched-chain starter units also include a putative methyltransferase domain (MT_L), which precedes the GNAT domain in the module (Figure 4.1a, b, c) [21, 45]. This path to incorporation of branched-chain starter units was identified recently in the annotation of gene clusters for biosynthesis of gephyronic acid (Gph) [21] and apratoxin A (Apr) [25, 184], a Sec61 inhibitor [26, 27]. Based on the natural product structures, the AprA initiation module installs a pivaloyl group, whereas

GphF introduces an isobutyryl group into gephyronic acid (Figure 4.1a, b). Isotopic labeling of the gephyronic acid producer *Cystobacter violaceus* with ^{13}C-methyl-L-methionine demonstrated that the isobutyryl starter group is derived from *S*-adenosylmethionine (SAM) and not from valine, likely via GphF MT_L [21]. The MT_L-containing initiation modules also include a mysterious ~150-residue N-terminal "adaptor region" (AR), which has no detectable homology to proteins of known structure or function. The AR domain of CurA (Figure 4.1f) enhanced, but was not essential to, the GNAT *S*-acyltransfer activity [31]. Pivaloyl-producing modules, such as AprA, include an additional methyltransferase ($MT2_L$) following the GNAT (Figure 4.1a) [25]. The MT_L sequences are highly conserved, but are distantly related to methyltransferases of known structure and to $MT2_L$.

Here we describe initial steps in the production of branched-chain polyketide starter units by GNAT initiation modules. Through structural and biochemical characterization of AprA MT_L we discovered an unusual mononuclear iron-dependent methyltransferase, which catalyzes two methyl transfer reactions to produce dimethylmalonyl (Me_2Mal-) ACP. MeMal- or Me_2Mal-ACP produced by MT_L homologs can be further decarboxylated by GNAT to yield propionyl or isobutyryl starter units. In pivalate-producing modules, such as AprA, the GNAT is truncated and not catalytically active. Characterization of AprA MT_L resolves the functional annotation of GNAT initiation modules from several biosynthetic pathways, including gephyronic acid [21], myxovirescin [97], saxitoxin [45] and bryostatin [20].

Experimental Procedures

Construct design

All primers are listed in Table 4.1. Full-length *aprA* was amplified from an apratoxin fosmid library [25] and cloned into pMCSG7 [166] via ligation independent cloning (LIC) to create pAPS1. The region encoding MT_L -ΨGNAT (AprA residues 2–629) was subcloned from pAPS1 to create pMAS286. This plasmid contains two PCR-induced missense mutations (S274I and Q528P), which were corrected using the QuickChange protocol (Stratagene) to create pMAS354. All *aprA* site-directed mutants were created from pMAS286. The region encoding AprA ΨGNAT (AprA residues 503–629) was subcloned from pAPS1 to create pMAS246. The region encoding AprA ACP (AprA residues 1058–1138) was subcloned from pAPS1 to create pAPS2. Plasmid pAPS3 was created by subcloning the sequence encoding GphF MT_L -GNAT (GphF residues 2–

696) from a partial *gphF* clone kindly provided by Richard Taylor (Notre Dame University) [21]. A GphF/H660A mutation was introduced (pMAS335) to block GphF GNAT decarboxylase activity. DNA sequencing was performed at the University of Michigan DNA Sequencing Core to verify all constructs and mutations.

Table 4.1 Primers for subcloning MT_L-ΨGNAT

AprA MT_L-ΨGNAT FWD	pMAS 286	5'-**TACTTCCAATCCAATGCC**CTAGATAAAATAAATCGTTATGCTCAT -3'
AprA MT_L-ΨGNAT REV		5'-**TTATCCACTTCCAATGCTA**CTTTCAAAAATCAAAGTGGTATCAAT CT-3'
AprA ΨGNAT FWD	pMAS 246	5'-**TACTTCCAATCCAATGCGCA**ATCTTCAGTTTTAAACAAAAAGCTT ATAG-3'
AprA ΨGNAT REV		5'- **TTATCCACTTCCAATGCTA**CTTTCAAAAATCAAAGTGGTATCAA TCT -3'
GphF MT_L-GNAT FWD	pAPS 3	5'-**TACTTCCAATCCAATGCC**CTTGAGCTACTCAACCAACACGC-3'
GphF MT_L-GNAT REV		5'-**TTATCCACTTCCAATGCTA**GGCCGCCCCACGGAGCG-3'
AprA ACP FWD	pAPS 2	5'-TACTTCCAATCCAATGCTGAAATTTTTGAACAGGAATGTC-3'
AprA ACP REV		5'-TTATCCACTTCCAATGTTAACTAAAATTAATATCTTCCCGTTCTG-3'
AprA MT_L-ΨGNAT I274S FWD	pMAS 354	5'-GCCAATAGAAGAGCAACCAAGCTATATAGTAGACATGGGTT-3'
AprA MT_L-ΨGNAT I274S REV		5'-AACCCATGTCTACTATATAGCTTGGTTGCTCTTCTATTGGC-3'
AprA MT_L-ΨGNAT P528Q FWD		5'-GAAGGGTTCGTTTACAGGTTGATTAAATGTCGCACACTTT-3'
AprA MT_L -ΨGNAT P528Q REV		5'-AAAGTGTGCGACATTTAATCAACCTGTAAACGAACCCTTC-3'
GphF H660A FWD	pMAS 335	5'-GGTGCTCGGGTTCGCCCTCGGTCACGGC-3'
GphF H660A REV		5'-GCCGTGACCGAGGGCGAACCCGAGCACC-3'
AprA MT_L-ΨGNAT R196E FWD	pMAS 362	5'-GACGCTGTTGTTCCTAAGTTTAAGGACTCGTCACGCATAAATCTAC CTATATCTG-3'
AprA MT_L-ΨGNAT R196E REV		5'-CAGATATAGGTAGATTTATGCGTGACGAGTCCTTAAACTTAGGAA CAACAGCGTC-3'
AprA MT_L-ΨGNAT Y206F FWD	pMAS 360	5'-CTTAGGAACAACAGCGTCTTTTGCTCCTATGTTGTTAC-3'
AprA MT_L-ΨGNAT Y206F REV		5'-GTAACAACATAGGAGCAAAAGACGCTGTTGTTCCTAAG-3'
AprA MT_L-ΨGNAT E233A FWD	pMAS 336	5'-AGTATTTCAAAGAAATAAGACTGAAAAAGCAAGACACGTAAATA GAACATTAAATGTAG-3'

AprA MT_L-ΨGNAT E233A REV	pMAS 336	5’-CTACATTTAATGTTCTATTTACGTGTCTTGCTTTTTCAGTCTTATTTCTTTGAAATACT-3’
AprA MT_L-ΨGNAT H235N FWD	pMAS 346	5’-CAGAGAGTATTTCAAAGAAATAAGACTGAAAAAGAAAGAAACGTAAATAGAACATTAAAT-3’
AprA MT_L-ΨGNAT H235N REV		5’-ATTTAATGTTCTATTTACGTTTCTTTCTTTTTCAGTCTTATTTCTTTGAAATACTCTCTG-3’
AprA MT_L-ΨGNAT R238A FWD	pMAS 357	5’-GAAATAAGACTGAAAAAGAAAGACACGTAAATGCAACATTAAATGTAGTAGCAAGTG-3’
AprA MT_L-ΨGNAT R238A REV		5’-CACTTGCTACTACATTTAATGTTGCATTTACGTGTCTTTCTTTTTCAGTCTTATTTC-3’
AprA MT_L-ΨGNAT N241A FWD	pMAS 359	5’-ACTGAAAAAGAAAGACACGTAAATAGAACATTAGCTGTAGTAGCAAGTGGCT-3’
AprA MT_L-ΨGNAT N241A REV		5’-AGCCACTTGCTACTACAGCTAATGTTCTATTTACGTGTCTTTCTTTTTCAGT-3’
AprA MT_L -ΨGNAT S245A FWD	pMAS 358	5’-AGACACGTAAATAGAACATTAAATGTAGTAGCAGCTGGCTTTCAACACGAAA-3’
AprA MT_L-ΨGNAT S245A REV		5’-TTTCGTGTTGAAAGCCAGCTGCTACTACATTTAATGTTCTATTTACGTGTCT-3’
AprA MT_L-ΨGNAT H249A FWD	pMAS 347	5’-TAAATGTAGTAGCAAGTGGCTTTCAAGCCGAAAAGTTTTTTGCCGATAC-3’
AprA MT_L-ΨGNAT H249A REV		5’-GTATCGGCAAAAAACTTTTCGGCTTGAAAGCCACTTGCTACTACATTTA-3’
AprA MT_L-ΨGNAT K251E FWD	pMAS 363	5’-GTATCGGCAAAAAACTCTTCGTGTTGAAAGCCACTTGCTA-3’
AprA MT_L-ΨGNAT K251E REV		5’-TAGCAAGTGGCTTTCAACACGAAGAGTTTTTTGCCGATAC-3’
AprA MT_L-ΨGNAT D370N FWD	pMAS 364	5’-GCTATAAATGGTCGGTTATGATCTAGGAATGAACGAATATGTAATACT-3’
AprA MT_L-ΨGNAT D370N REV		5’-AGTATTACATATTCGTTCATTCCTAGATCATAACCGACCATTTATAGC-3’
AprA MT_L-ΨGNAT Y455F FWD	pMAS 361	5’-GTGAATCTCTGCATTTTGATGCCTTTCATGCCTTCTCAAT-3’
AprA MT_L-ΨGNAT Y455F REV		5’-ATTGAGAAGGCATGAAAGGCATCAAAATGCAGAGATTCAC-3’
AprA MT_L-ΨGNAT H462A FWD	pMAS 350	5’-CTATCATGCCTTCTCAATGCAAGCTTTAGTAGAAGCAGATGTGTTC-3’

AprA MT_L-ΨGNAT H462A REV	pMAS 350	5'-GAACACATCTGCTTCTACTAAAGCTTGCATTGAGAAGGCATGATA G-3'
AprA MT_L-ΨGNAT K490E FWD	pMAS 365	5'-GATTCGAGTCAATGGCAACGTCTCCGGATACTTACGAAAAGCTTC-3'
AprA MT_L-ΨGNAT K490E REV		5'-GAAGCTTTTCGTAAGTATCCGGAGACGTTGCCATTGACTCGAATC-3'
AprA MT_L-ΨGNAT R496A FWD	pMAS 351	5'-GTATCCGAAAACGTTGCCATTGACTGCAATCACAGTTAATCATTTT GAAAAG-3'
AprA MT_L-ΨGNAT R496A REV		5'-CTTTTCAAAATGATTAACTGTGATTGCAGTCAATGGCAACGTTTTC GGATAC-3'

Bold font indicates handles for ligation-independent insertion into expression vectors

Protein expression and purification

All plasmids were transformed into *Escherichia coli* strain BL21(DE3), grown in 0.5 L of TB media with 100 μg mL^{-1} ampicillin to an OD_{600} = 1–2 at 37°C, cooled to 20°C for 1 hr, induced with 200 μM IPTG and expressed overnight. 2L SelenoMet medium (AthenaES) containing 150 μg mL^{-1} of seleno-DL-methionine was used to produce selenomethionine (SeMet) labeled protein. Cells were grown to an OD_{600} = 0.6 at 37°C, cooled for 1 hr at 20°C, induced with 200 μM IPTG, and expressed overnight. To produce holo-ACP pAPS2 was transformed into *Escherichia coli* strain BAP1[185] and grown, induced and expressed as above.

AprA MT_L-ΨGNAT, AprA ΨGNAT, and GphF MT_L-GNAT cell pellets were resuspended in Tris buffer A (50 mM Tris pH 7.4, 300 mM NaCl, 10% (v/v) glycerol, 20 mM imidazole) with 0.1 mg mL^{-1} lysozyme, 0.05 mg mL^{-1} DNase, and 2 mM $MgCl_2$. Resuspended cell pellets were incubated on ice for 30 min, lysed by sonication, and cleared by centrifugation (38,650 x g, 30 min). Filtered supernatant was loaded onto a 5 mL His trap column (GE Healthcare), washed with 10 column volumes of Tris buffer A, and eluted with a 20–400 mM linear gradient of imidazole (Tris buffer B) over 10 column volumes. Proteins were further purified by gel filtration chromatography (HiLoad 16/60 Superdex S200) with Tris buffer C (50 mM Tris pH 7.4, 150 mM NaCl, 10% (v/v) glycerol). ~80 mg L^{-1} of MT_L -GNAT was obtained.

AprA ACP cell pellets were resuspended in ACP buffer A (100 mM Tris pH 7.4, 500 mM NaCl, 10% (v/v) glycerol, 5 mM TCEP, 20 mM imidazole) with 0.1 mg mL^{-1} lysozyme, 0.05 mg mL^{-1} DNase, and 2 mM $MgCl_2$, and treated identically to MT_L-GNAT cell pellets. ACP buffer B

(100 mM Tris pH 7.4, 500 mM NaCl, 10% (v/v) glycerol, 5 mM TCEP, 300 mM imidazole) was used to elute AprA ACP from a 5 mL His trap column over 10 column volumes. AprA ACP was further purified using gel filtration chromatography (HiLoad 16/60 Superdex S75) with ACP buffer C (100 mM Tris pH 7.4, 250 mM NaCl, 5% (v/v) glycerol, 5 mM TCEP). ~50 mg L^{-1} of AprA ACP was obtained. CurJ *C*-MT and JamJ ACP were expressed and purified as described previously [117].

Production of acyl-ACPs

180 μM AprA ACP was incubated with 20 μM *Streptomyces verticillus* phosphopantetheinyl transferase (SVP) [142], 20 mM $MgCl_2$, and 3–4 fold molar excess of acetyl-, isobutyryl-, Mal-, MeMal-, or propionyl-CoA for 4 hr at 30°C in Tris Buffer C. Acyl-ACPs were purified from reaction mixtures using gel filtration chromatography (HiLoad 16/60 Superdex S75) with ACP buffer C (100 mM Tris pH 7.4, 250 mM NaCl, 5% (v/v) glycerol, 5 mM TCEP).

Enzyme assays

All assays were conducted in triplicate. 10 μL AprA methylation reaction mixtures containing 100 μM AprA acetyl-, isobutyryl-, Mal-, MeMal-, or propionyl-ACP were incubated with 25 μM wild type or mutagenized AprA MT_L-ΨGNAT or AprA ΨGNAT, 1.35 mM SAM, and 0.5 mM $CaCl_2$, $CoCl_2$, $CuSO_4$, $(NH_4)_2Fe(SO_4)_2$, $NiSO_4$, $MgCl_2$, $MnCl_2$, Na_2MoO_4, or $ZnCl_2$, in 50 mM Hepes 7.4, 150 mM NaCl. For anaerobic reactions, all buffers were degassed through several freeze-pump-thaw cycles and brought into an anaerobic glove box. Solid $(NH_4)_2Fe(SO_4)_2$ and SAM, and aliquots of AprA MT_L-ΨGNAT and ACP were preincubated for 1 hr in the anaerobic environment in a 0–4°C CoolBox prior to preparing Fe^{2+} and SAM solutions and reaction mixtures. Reactions were incubated 5 hr at 30°C (or 4 hr for the time course). AprA acyltransfer reaction mixtures (10 μL) containing 100 μM AprA holo-ACP were incubated with 25 μM AprA ΨGNAT and 0.5 mM Mal- or MeMal-CoA in 50 mM Hepes 7.4, 150 mM NaCl. Reactions were incubated at 30°C for 10 hr. CurJ *C*-MT reaction mixtures (10 μL) containing 100 μM JamJ apo ACP, 25 μM CurJ *C*-MT, 10 μM SVP, 1 mM $MgCl_2$, 1.35 mM SAM, 0.5 mM $(NH_4)_2Fe(SO_4)_2$, and 0.5 mM Mal-CoA in 50 mM Hepes 7.4, 150 mM NaCl. Reactions were incubated at 30°C for 5 hr. GphF reaction mixtures (10 μL) contained 100 μM AprA Mal-ACP, 50 μM GphF MT_L-GNAT H660A, 1.35 mM SAM, and 0.5 mM $(NH_4)_2Fe(SO_4)_2$ or $MnCl_2$, in 50 mM Hepes pH 7.4, 150 mM NaCl. Reactions were incubated at 20°C for 24 hr and quenched with

10% (v/v) formic acid prior to LC/MS analysis. Enzyme assays were slightly modified to assess AprA MT_L -ΨGNAT activity on CoA substrates by monitoring absorbance at 254 nm of reaction mixtures separated by HPLC.

LC/MS analysis

Assay reaction mixtures (0.1 μL) were separated by reverse phase HPLC (Phenomenex Aeris widepore C4 column 3.6 μM, 50 x 2.10 mm) at a flow rate of 0.5 mL min^{-1} in H_2O with 0.2% (v/v) formic acid. Protein was eluted using a gradient of 5-100% acetonitrile with 0.2% (v/v) formic acid over 4 minutes. LC-MS analysis (Agilent Q-TOF 6545) using a simultaneous intact protein and phosphopantetheine (Ppant) ejection method [150, 169] was carried out under the following conditions: fragmentor voltage, 300 V; skimmer voltage, 75 V; nozzle voltage, 1000 V; sheath gas temperature, 350 °C; drying gas temperature, 325 °C.

MassHunter Qualitative Analysis Software (Agilent) was used for LC-MS data analysis. Intact protein masses were obtained using the maximum entropy deconvolution algorithm. Significant proportions of the Mal-, MeMal-, and Me_2Mal-Ppant ejection fragments underwent in-source decay resulting in acetyl-, propionyl-, and isobuturyl-Ppant fragments. As lower instrument voltages decreased in-source decay and also sensitivity, the abundances of related Ppant fragments were summed (acetyl + Mal, propionyl + MeMal, isobutyryl + Me_2Mal-) to calculate the relative abundance of Mal-, MeMal-, and Me_2Mal-ACP. Data were excluded for holo ACP, which resulted from the spontaneous hydrolysis of Mal-, MeMal-, or Me_2Mal-ACP.

HPLC assays

25 μL reaction mixtures containing 125 μM AprA MT_L-ΨGNAT, 500 μM Mal-CoA or Mal-ACP, 6.75 mM SAM, 2.5 mM $(NH_4)_2Fe(SO_4)_2$, 50 mM Hepes 7.4, 150 mM NaCl were incubated for 5 hr at 30°C and quenched with the addition of 50 μL methanol. Precipitated protein was removed by centrifugation. To detect CoA products, end-point reaction mixtures (10 μL) containing 75 μM AprA MT_L-ΨGNAT, 2 mM SAM, 0.5 mM $(NH_4)_2Fe(SO_4)_2$, 1.5 mM Mal-CoA or MeMal-CoA, 50 mM Hepes 7.4, 150 mM NaCl were incubated 12 hr at 30°C and quenched with the addition of 20 μL methanol.

Small molecule components of reaction mixtures (10 μL) were separated by reverse phase HPLC (Phenomenex Luna C18 column 5 μM, 250 x 4.6 mm) following a previously reported method[186]; absorbance was monitored at 254 nm. SAH production was quantified based on

integrated peak areas of SAH compared to a SAH standard curve. For reactions on Mal-ACP, the LC-MS and HPLC assays resulted in nearly identical levels of activity after correction for assay differences in enzyme and substrate concentrations (108 +/- 6 uM SAH produced via LC-MS, 106 +/- 15 uM SAH produced via HPLC).

Protein crystallization and structure determination

Crystal structures of AprA MT_L-ΨGNAT were solved for a serendipitous double-substitution (S274I/Q528P) that was introduced during amplification of the gene and was essential to crystallization. The double substitution had no effect on catalytic activity. Native and SeMet AprA MT_L -ΨGNAT/S274I/Q528P was crystallized by vapor diffusion in a 1:2 µL mixture of protein stock (10–11 mg mL^{-1} AprA MT_L-ΨGNAT in Tris buffer C with 1 mM SAM) and well solution (2.4–2.5 M $(NH_4)_2SO_4$, 0.1 M Tris pH 8.5) at 4°C. Crystals appeared overnight and continued to grow over several days. Crystals were cryoprotected in 2.4 M $(NH_4)_2SO_4$ supplemented with 25% (v/v) glycerol and flash cooled in liquid nitrogen. Diffraction data for all structures were processed using XDS (Tables 4.2-4.6) [128]. The SeMet AprA MT_L-ΨGNAT structure was solved by single-wavelength anomalous diffraction (Se SAD) using Phenix AutoSol in the Phenix Software suite [130, 170]. A nearly complete model was built using Phenix AutoBuild [171]. Crystals of the native and SeMet MT_L -ΨGNAT/S274I/Q528P were isomorphous. All refined structures were completed by iterative rounds of model building and refinement using Coot [132] and Phenix.refine with three translation/liberation/screw groups corresponding to the MT_L lid, MT_L core and ΨGNAT. Electron density is complete for all but the 14 N-terminal amino acids of the His tag, a 13-residue region between the MT core and lid (residues 227–241), and the three C-terminal residues.

Mn-bound AprA MT_L-ΨGNAT/S274I/Q528P was crystallized by vapor diffusion in a 2:1 µL mixture of protein stock (10-11 mg mL^{-1} of AprA MT_L-ΨGNAT in Tris buffer C with 1 mM SAM and 5 mM $MnCl_2$) and well solution (0.15 M DL-malic acid pH 7.0, 20% (v/v) PEG 3350) at 20°C. Plate-like crystals appeared overnight. Crystals were cryoprotected with well solution supplemented with 15% (v/v) glycerol and flash cooled in liquid nitrogen. A fully refined apo AprA MT_L-GNAT structure was used to solve the structure of Mn bound AprA MT_L-ΨGNAT by molecular replacement using Phaser in the Phenix Software Suite [172]. Electron density is complete

for all residues but the 16 N-terminal residues of the His-tag and 5 residues of the lid (residues 148–156).

Malonate- and Mn-bound AprA MT_L-ΨGNAT/S274I/Q528P was crystallized by vapor diffusion in a 1:1 µL mixture of protein stock (10–11 mg mL^{-1} AprA MT_L -ΨGNAT in Tris buffer C with 1 mM SAM and 5 mM $MnCl_2$) and well solution (0.05 M sodium malonate, 16% (v/v) PEG 3350) at 4°C. Microseeding was used to obtain single crystals. Crystals were cryoprotected with well solution supplemented with 15% (v/v) glycerol and flash cooled in liquid nitrogen. Crystals were isomorphous with those of Mn-bound MT_L-ΨGNAT/S274I/Q528P. Electron density is missing for 23 N-terminal residues of the His-tag and five amino acids of the MT_L lid (residues 150–154).

All structures were validated using MolProbity (Figures 4.2-4.4, Table 4.2) [133].Figures were prepared in PyMol [136]. The AprA ACP homology model was produced using I-TASSER [139-141]. Electrostatic potential surfaces were calculated in APBS with default settings [187]. Sequence alignments were created using Clustal in Jalview [137, 138]

Table 4.2 AprA MT_L-ΨGNAT crystallographic information

Data Collection	SeMet Metal Free	Metal Free	Mn	Mn + Malonate
Space group	*P*4$_2$22	*P*4$_2$22	*P*2$_1$2$_1$2$_1$	*P*2$_1$2$_1$2$_1$
Unit cell, a,b,c (Å)	152.9, 152.9, 95.0	152.8, 152.8, 95.0	60.5, 87.1, 136.0	60.7, 88.4, 136.8
X-ray source	APS 23ID-B	APS 23ID-B	APS 23-IDB	APS 23ID-D
Wavelength (Å)	0.979	1.033	1.033	1.033
d_{min} (Å)	2.48 (2.57-2.48)[a]	2.39 (2.48-2.39)	1.78 (1.85-1.78)	1.85 (1.91-1.85)
R-merge	0.1864 (1.98)	0.1052 (1.96)	0.0715 (0.910)	0.0987 (0.8719)
Avg I/σ(I)	8.96 (0.89)	11.78 (0.71)	10.72 (1.17)	7.72 (1.05)
Completeness (%)	99.7 (97.8)	93.1 (91.0)	95.3 (93.2)	95.9 (95.6)
Multiplicity	11.8 (12.0)	7.2 (7.3)	3.5 (3.2)	3.4 (3.3)
Total observations	474,533 (46,720)	305,285 (30,710)	227,183 (20,386)	211,330 (19,660)
Wilson B factor (Å^2)	60.61	64.26	27.54	26.21
$CC_{1/2}$	0.996 (0.549)	0.998 (0.289)	0.998 (0.671)	0.995 (0.494)
CC*	0.999 (0.842)	1.00 (0.670)	0.999 (0.896)	0.999 (0.813)
Refinement				
Data range (Å)		47.0–2.39	45.53-1.78	45.4-1.85
Reflections used in refinement		41,996	65,829	61,416
R_{work}/R_{free} (%)		18.5/22.4	18.1/22.5	18.5/21.5
Number of non-hydrogen atoms		5,309	5,623	5,516
protein		5,074	5,094	5,056
ligands		72	7	14
water		163	522	446
Amino acid residues		622	631	624
Deviation from ideality				
bond lengths (Å)		0.008	0.007	0.006
bond angles (°)		0.99	0.94	0.84
Average B-factor (Å^2)		74.8	38.5	37.5
protein		74.1	37.5	37.3
ligands		131.7	60.9	43.8
solvent		74.5	48.3	40.2
Ramachandran plot				
favored (%)		95.2	97.1	97.7
allowed (%)		4.2	2.6	2.3
outliers (%)		0.6	0.3	0
PDB ID		6B39	6B3A	6B3B

[a]values in parentheses designate outer shell

Table 4.3 Scaling statistics for AprA MT_L-ΨGNAT SeMet metal free

Resolution Limit (Å)	Number of Reflections						
	Observed	Unique	Possible	Completeness	R-Factor	I/σ	$CC_{1/2}$
11.11	10422	847	860	98.50%	8.50%	26.38	99.6
7.86	20608	1555	1555	100.00%	8.70%	26.5	99.5
6.41	25355	2006	2006	100.00%	10.20%	21.51	99.6
5.56	29923	2398	2398	100.00%	11.40%	19.22	99.5
4.97	30996	2682	2684	99.90%	11.00%	18.45	99.5
4.54	37258	2994	2994	100.00%	11.20%	20.04	99.5
4.2	41531	3242	3242	100.00%	11.40%	19.3	99.5
3.93	45147	3468	3468	100.00%	13.30%	17.15	99.4
3.7	49005	3726	3726	100.00%	14.80%	15.29	99.4
3.51	51644	3901	3901	100.00%	17.50%	13.17	99.2
3.35	54991	4135	4135	100.00%	21.10%	11.11	99.0
3.21	54532	4298	4298	100.00%	27.00%	8.47	98.6
3.08	54158	4502	4502	100.00%	36.20%	6.45	97.0
2.97	59008	4668	4668	100.00%	50.00%	4.77	95.4
2.87	56340	4848	4848	100.00%	64.50%	3.53	93.0
2.78	59807	4979	4980	100.00%	81.60%	2.85	87.6
2.69	64445	5189	5189	100.00%	99.60%	2.3	84.4
2.62	67070	5327	5327	100.00%	121.10%	1.84	78.9
2.55	69317	5456	5457	100.00%	168.90%	1.27	64.0
2.48	66781	5466	5590	97.80%	194.50%	1	55.2
Total	948338	75687	75828	99.80%	17.40%	9.14	99.6

Table 4.4 Scaling statistics for AprA MT_L-ΨGNAT metal free (PDB ID 6B39)

Resolution Limit (Å)	Number of Reflections						
	Observed	Unique	Possible	Completeness	R-Factor	I/σ	$CC_{1/2}$
7.11	12161	1690	1916	88.20%	3.10%	49.68	99.9
5.05	19799	2866	3173	90.30%	5.40%	30.63	99.8
4.13	26980	3689	4001	92.20%	5.70%	30.95	99.8
3.58	33210	4371	4713	92.70%	8.20%	21.5	99.6
3.21	35469	4939	5288	93.40%	13.80%	11.69	99.1
2.93	37846	5467	5821	93.90%	31.00%	5.26	95.2
2.71	42800	5928	6275	94.50%	60.50%	2.7	85.1
2.54	47067	6424	6747	95.20%	104.10%	1.49	62.9
2.39	49976	6846	7139	95.90%	176.80%	0.8	34.9
Total	305308	42220	45073	93.70%	10.50%	11.78	99.8

Table 4.5 Scaling statistics for AprA MT_L-ΨGNAT Mn bound (PDB ID 6B3A)

Resolution Limit (Å)	Number of Reflections						
	Observed	Unique	Possible	Completeness	R-Factor	I/σ	$CC_{1/2}$
5.32	9053	2461	2827	87.10%	3.20%	32.02	99.9
3.78	15276	4343	4814	90.20%	3.90%	29.88	99.7
3.09	20937	5705	6156	92.70%	5.00%	24.13	99.6
2.67	24150	6803	7189	94.60%	7.10%	16.12	99.4
2.39	25929	7761	8124	95.50%	10.20%	11.05	98.8
2.18	29736	8605	8923	96.40%	14.70%	8.15	97.9
2.02	33185	9454	9680	97.70%	22.50%	5.34	96.2
1.89	34392	10175	10333	98.50%	39.10%	2.88	90.3
1.78	34529	10533	11047	95.30%	79.10%	1.36	72.3
Total	227187	65840	69093	95.30%	7.10%	10.72	99.8

Table 4.6 Scaling statistics for AprA MT_L-ΨGNAT Mn and malonate bound (PDB ID 6B3B)

Resolution Limit (Å)	Number of Reflections						
	Observed	Unique	Possible	Completeness	R-Factor	I/σ	$CC_{1/2}$
5.5	7911	2328	2619	88.90%	4.70%	22.93	99.6
3.9	15229	4168	4486	92.90%	4.90%	22.44	99.5
3.19	17981	5313	5672	93.70%	6.00%	17.19	99.4
2.77	22808	6344	6653	95.40%	8.90%	11.64	98.9
2.47	25531	7325	7553	97.00%	13.30%	7.69	97.6
2.26	27133	7966	8259	96.50%	19.70%	5.25	94.8
2.09	30720	8763	8961	97.80%	28.60%	3.62	90.5
1.96	32109	9443	9611	98.30%	47.00%	2.15	77.4
1.85	31914	9777	10209	95.80%	75.50%	1.23	56.8
Total	211336	61427	64023	95.90%	9.90%	7.72	99.5

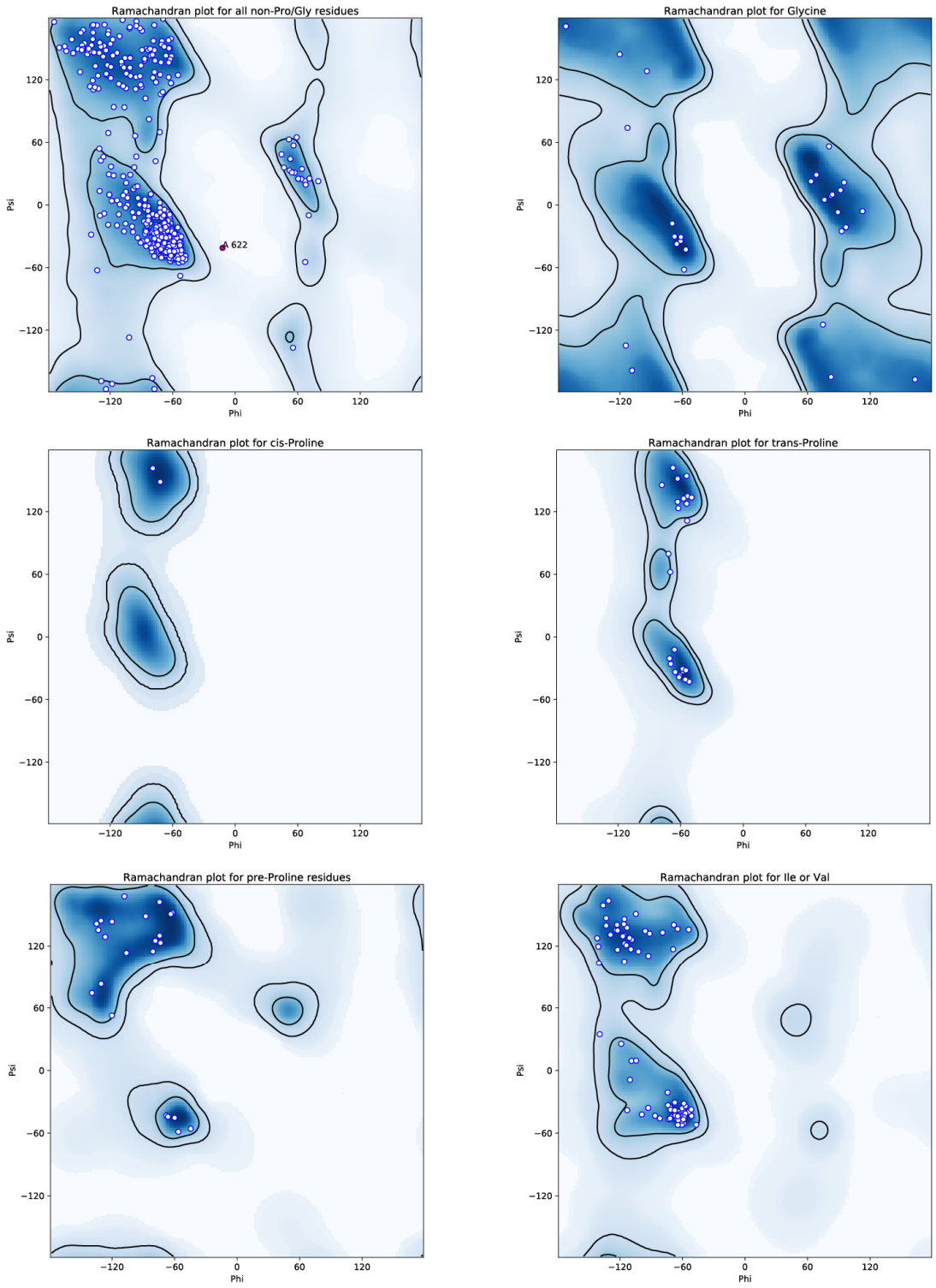


Figure 4.2 Ramachandran plots for AprA MT_L-ΨGNAT metal free
Ramachandran plots of final refined model of metal free AprA MT_L-ΨGNAT (PDB ID 6B39). Plots were generated using MolProbity[133].

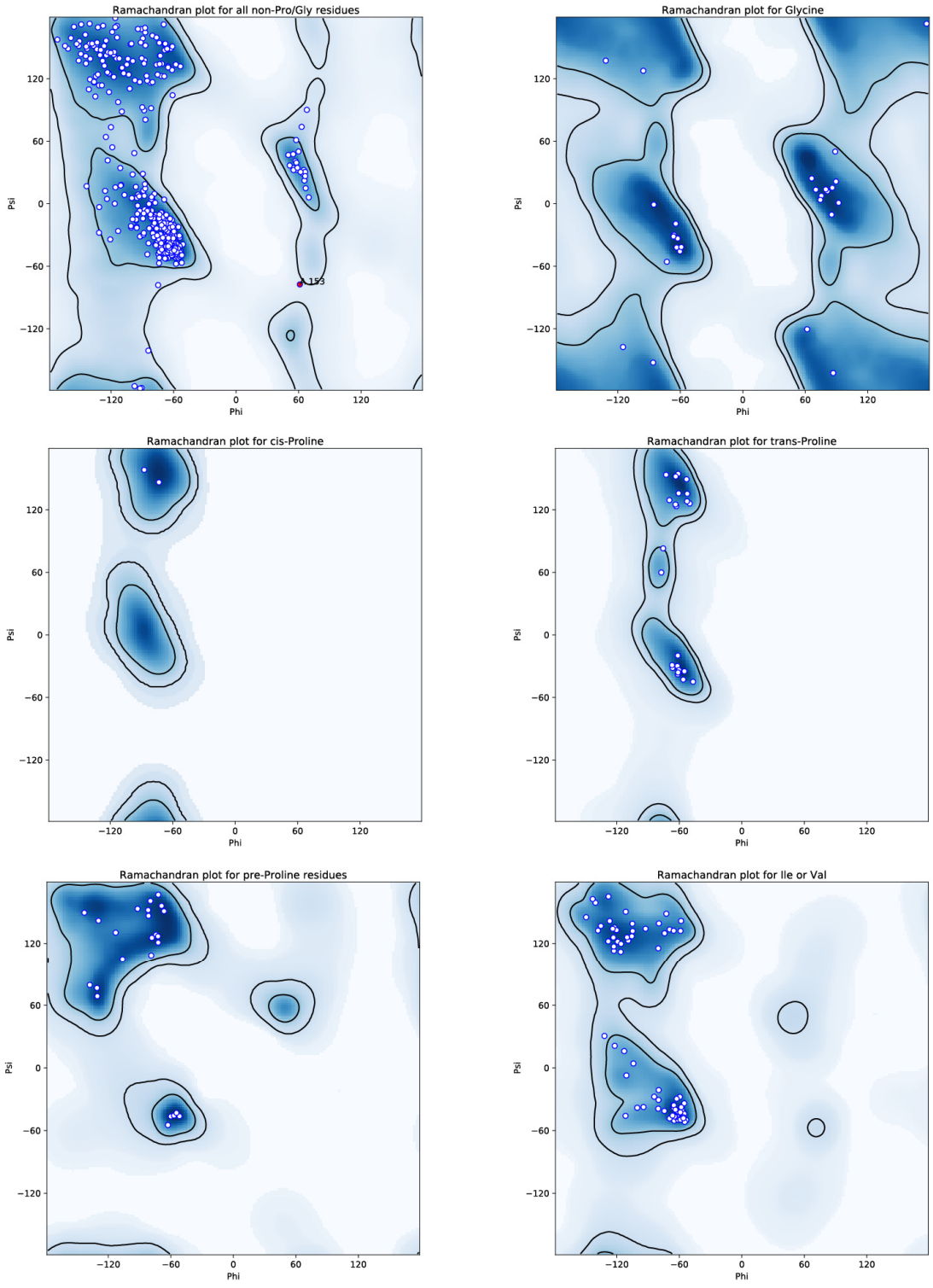


Figure 4.3 Ramachandran plots for Mn bound AprA MT_L-ΨGNAT
Ramachandran plots of final refined model of Mn bound MT_L-ΨGNAT (PDB ID 6B3A). Plots were generated using MolProbity[133]

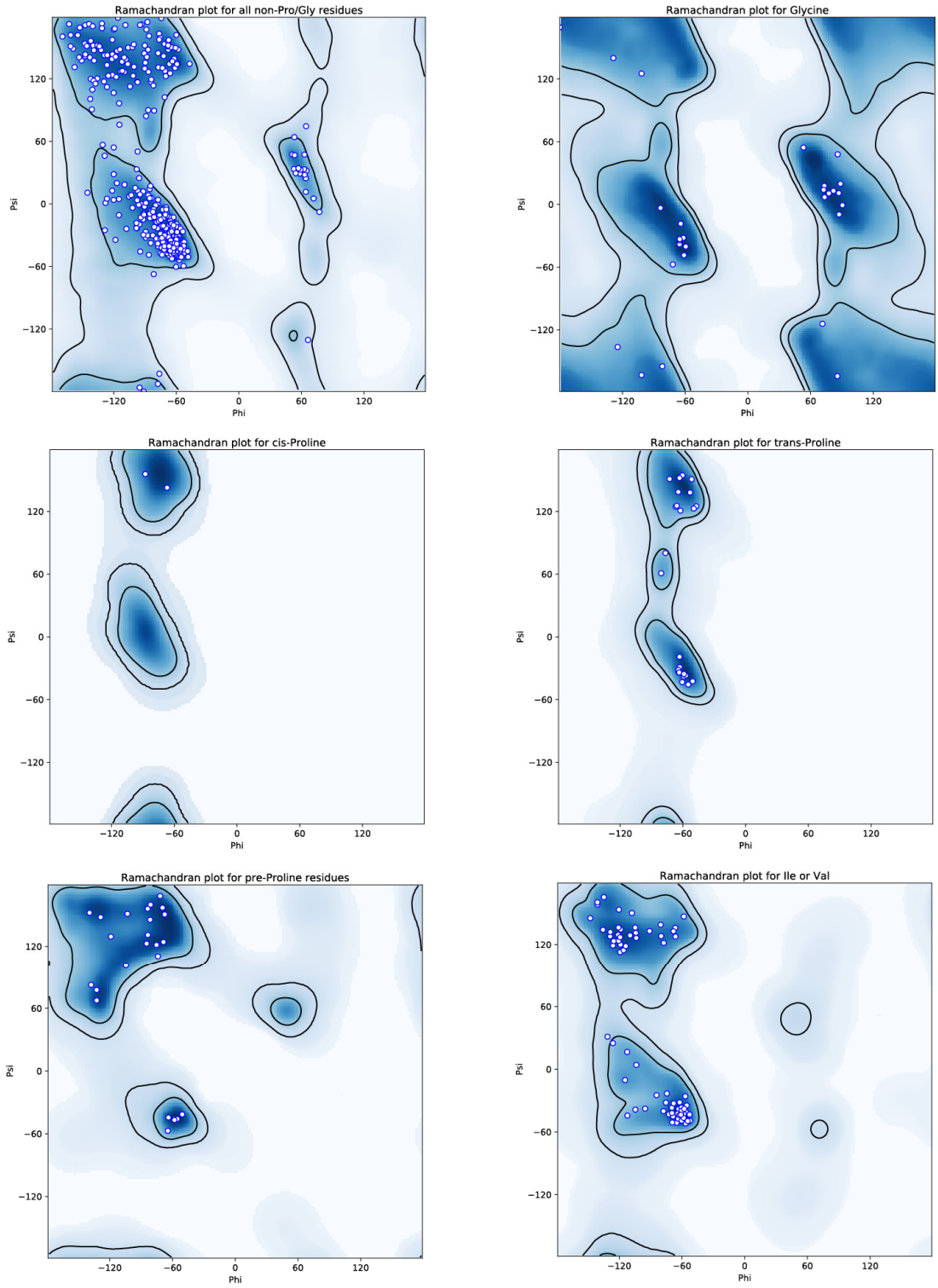


Figure 4.4 Ramachandran plots for Mn and malonate bound AprA MT_L-ΨGNAT
Ramachandran plots of final refined model of Mn bound MT_L-ΨGNAT in complex with malonate (PDB ID 6B3B). Plots were generated using MolProbity[133]

Results and Discussion

Bioinformatic analysis of MT_L-GNAT initiation modules

The AprA amino acid sequence was used to identify several homologs of AR-MT_L-GNAT in annotated and unannotated secondary metabolite biosynthetic pathways in proteobacteria, cyanobacteria, dinoflagellates, firmicutes, and planctomycetes. The homologous sequences from pathways with characterized natural products represent initiation modules that incorporate starter units larger than acetate (propionate, isobutyrate), which could be derived from the decarboxylation of MeMal- or Me_2Mal-CoA/ACP or from methylation of acetyl- or propionyl-CoA/ACP (Figure 4.1a-c). Our search identified an unannotated MT domain in the myxovirescin biosynthetic pathway where TaI module 1 contains an AR-MT_L-GNAT tridomain (Figure 4.1c) that could generate a propionyl starter unit for synthesis of the myxovirescin C3-hydroxyvaleryl moiety, consistent with ^{13}C labeling studies [97]. Another variant occurs in the rhizoxin pathway, which incorporates an unbranched acetyl starter unit despite the AR-MT_L-GNAT initiation module architecture (Figure 4.1d) [98]. However, the RhiA MT_L sequence lacks the characteristic SAM-binding motifs, indicating that it is nonfunctional (Figure 4.5). We also identified several AR-MT_L-GNAT initiation module variants that lack AR, MT_L or both [MT_L-GNAT-ACP (Figure 4.1e), AR-GNAT-ACP (Figure 4.1f), GNAT-ACP (Figure 4.1g)] and have predicted acetyl starter units. We conclude from the sequence analysis that AR-MT_L-GNAT initiation modules produce starter units larger than acetyl by SAM-dependent methyl transfer via the MT_L, and that both AR and a functional MT_L are required for the biosynthesis of *C*-methylated acetyl starter units.

AprA, the pivaloyl-producing initiation module in apratoxin biosynthesis, contains a second MT ($MT2_L$) following the GNAT (Figure 4.1a) [25]. Among several homologous sequences having both MT_L and $MT2_L$ domains, only AprA and BryX are from pathways with natural products of known structure. The *bryX* gene is a poorly characterized element of the gene cluster for bryostatin [20], a protein kinase C modulator [188]. The AR-MT_L-GNAT-$MT2_L$-ACP domain architecture of BryX is consistent with several isolated bryostatins that contain pivaloyl groups [20], leading to the prediction that $MT2_L$ is needed for a third methyl transfer in both apratoxin and bryostatin biosynthesis.

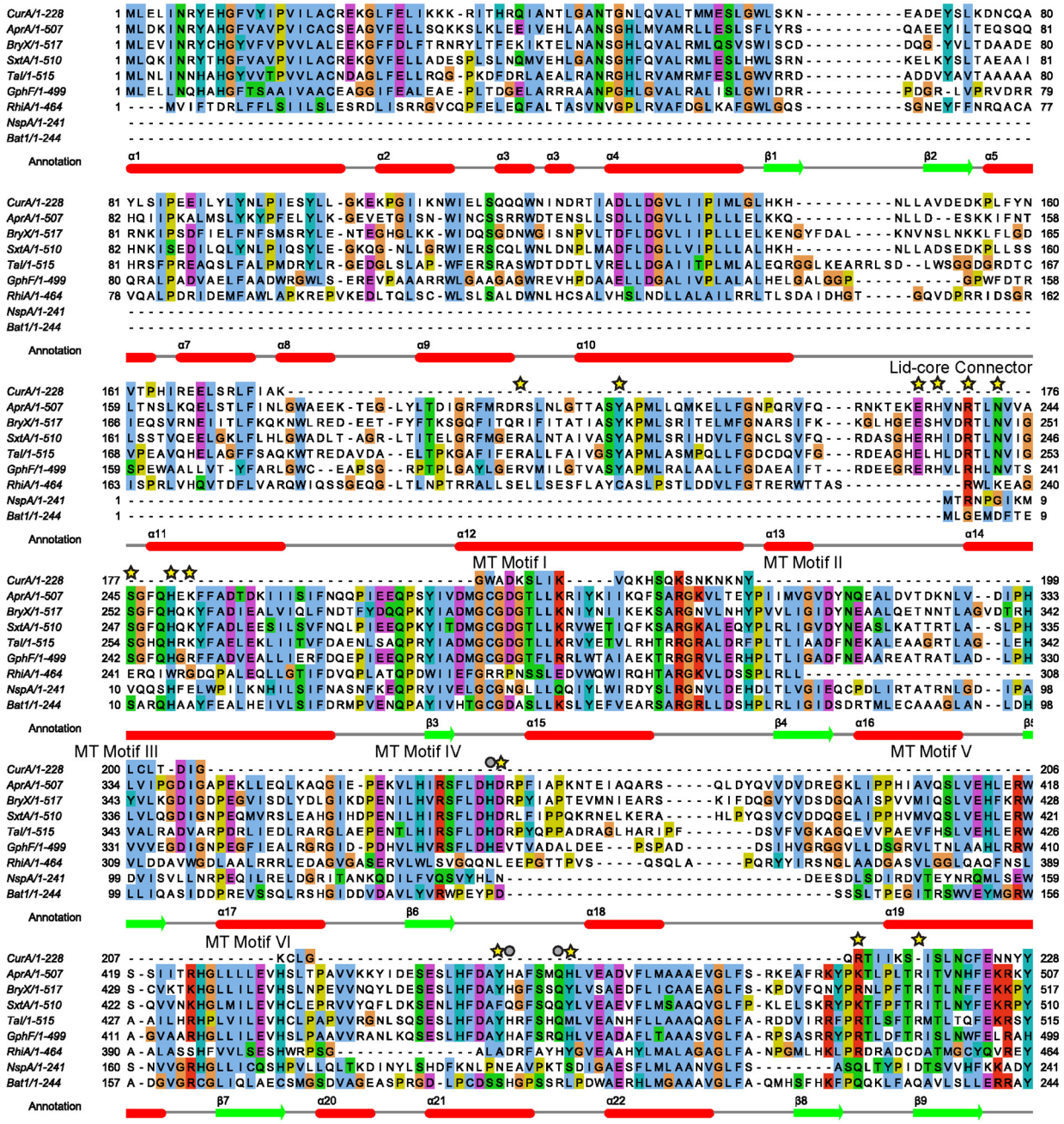


Figure 4.5 AR-MT1 sequence alignment with AprA MT_L secondary structure annotation
Metal ligands are indicated with gray spheres, and AprA amino acids subjected to mutagenesis are starred. Pathway abbreviations (GenBank accession codes) are as follows: Cur- curacin A (AEE88289.1), Apr- apratoxin A (WP_075900460), BryX- bryostatin (ABK51302.1), SxtA- saxitoxin (WP_009343302.1), TaI- myxovirescin A (WP_011553948.1), GphF- gephyronic acid (KF479198.1), RhiA- rhizoxin (WP_013435483.1), NspA- nosperin (ADA69237.1), Bat1- batumin (WP_052451043.1).

AprA AR-MT$_L$-GNAT structure

In order to understand the function of the common AR-MT$_L$-GNAT initiation module, we sought a stable AprA fragment for biochemical and structural studies. By screening several fragments for solubility, we identified a stable tridomain (residues 2–629) and solved a 2.4-Å crystal structure by single-wavelength anomalous diffraction (SAD) from the selenomethionyl protein (Figure 4.6a, Table 4.2). Crystal growth required two serendipitous amino acid substitutions (S274I, Q528P) that were introduced during cloning and are located in crystal contacts. AprA AR-MT$_L$-GNAT is monomeric in crystals and in solution as determined by gel filtration chromatography. The MT$_L$ domain (residues 241–505) has a typical class I MT fold and forms the SAM-binding core. The crystallization solution included SAM, but crystals contained the hydrolysis product *S*-adenosylhomocysteine (SAH) with clear density for the adenosine portion and poorly ordered homocysteine (Figure 4.7a, Figure 4.8a). The mysterious AR

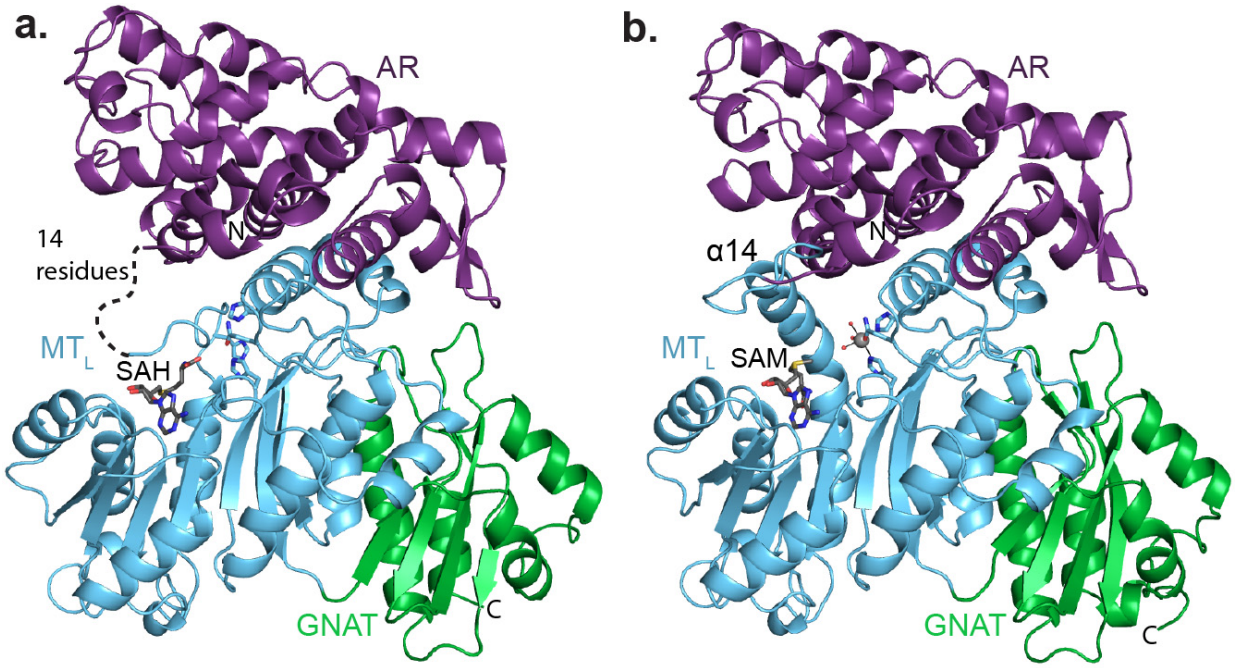


Figure 4.6 Structures of AprA AR-MT$_L$-ΨGNAT colored by structural region
(AR, purple; MT$_L$, blue; ΨGNAT, green). SAH or SAM is shown in sticks with atomic colors (C, gray; O, red; N, blue; S, yellow). Disordered residues in the lid-core connector (228–240) are indicated with a dashed line. **a.** AprA AR-MT$_L$-ΨGNAT. **b.** AprA AR-MT$_L$-ΨGNAT with bound Mn^{2+}. Mn is shown as a gray sphere and water ligands as red spheres.

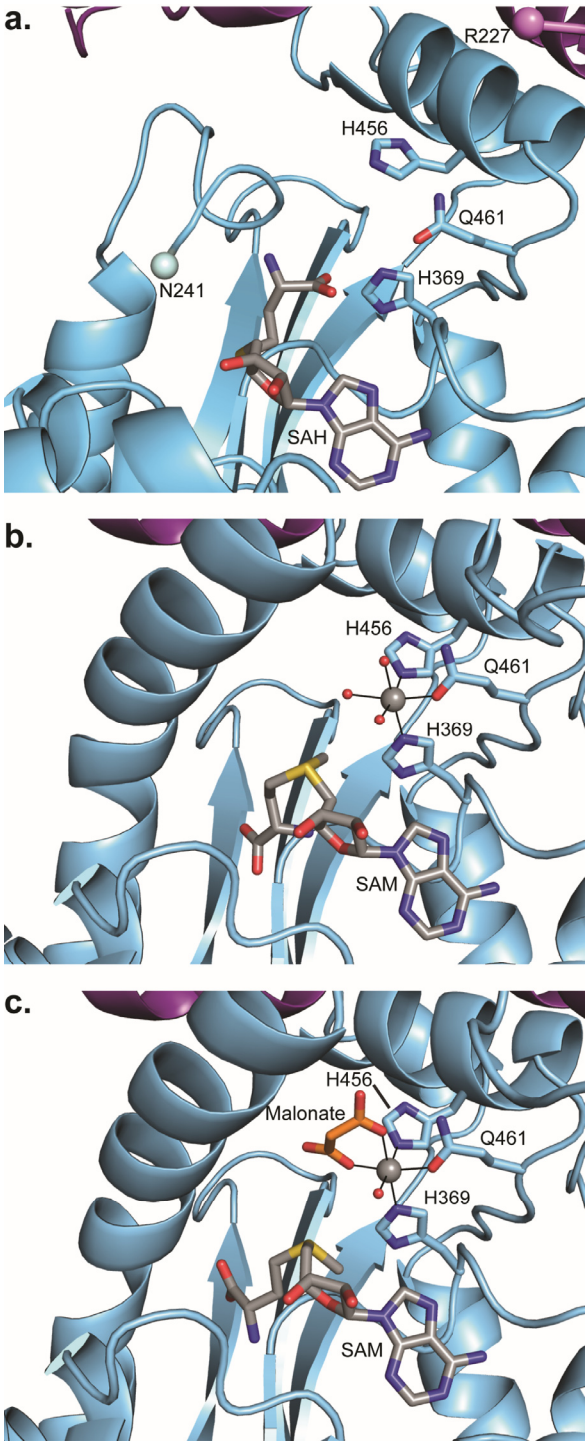


Figure 4.7 AprA MT_L active sites
a. Metal-free AprA MT_L-ΨGNAT with SAH. Arg227 before and Asn241 after the disordered lid-core linker are designated with spheres. **b.** AprA MT_L-ΨGNAT with Mn^{2+} and SAM. **c.** AprA MT_L-ΨGNAT substrate complex with Mn^{2+}, malonate and SAM. MT_L structural regions are colored as in Figure 4.6. Mn^{2+} (gray) and water ligands (red) are shown as spheres. Mn^{2+} ligands (blue C), SAH/SAM (gray C), and malonate (orange C) are shown in sticks with atomic colors.

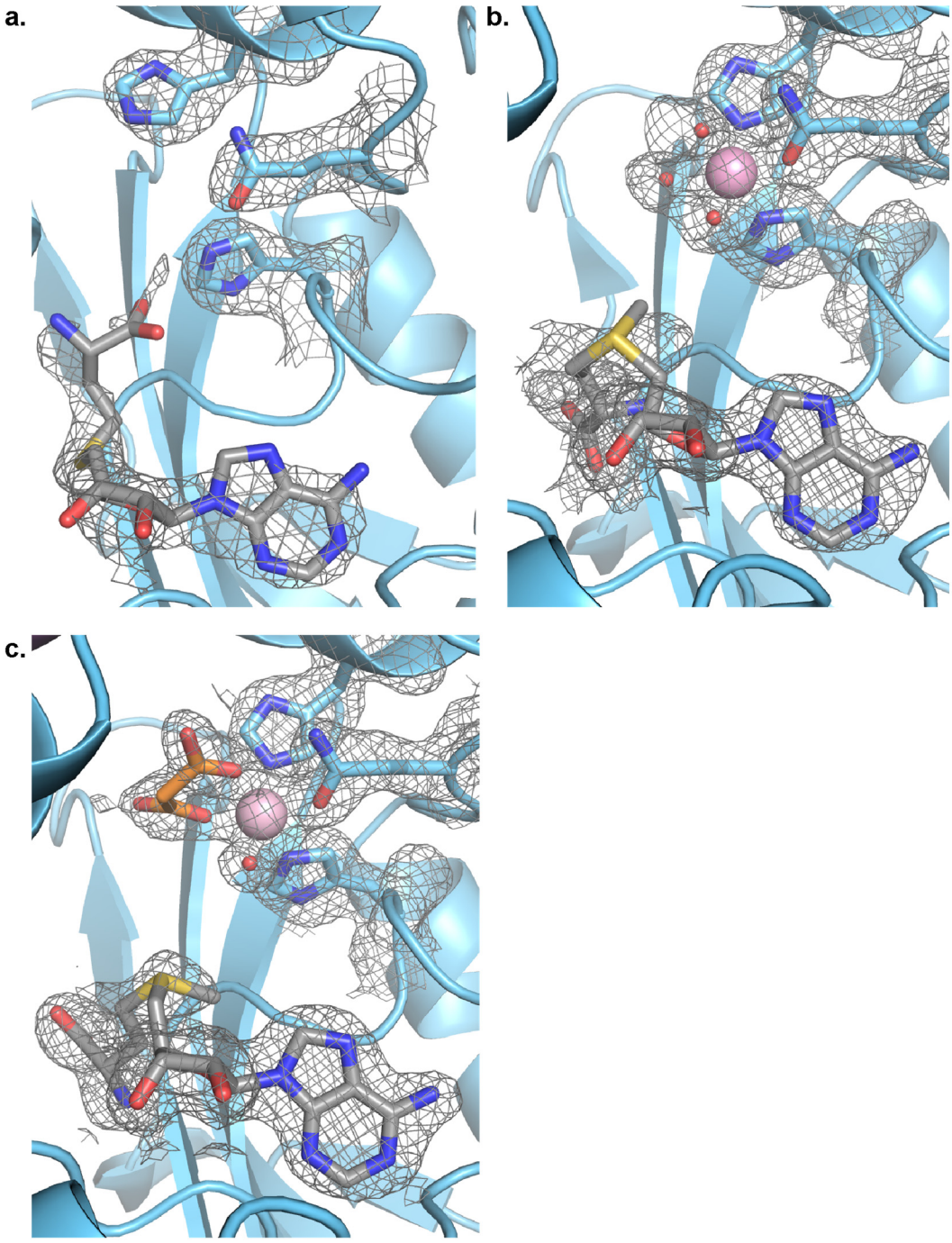


Figure 4.8 Omit density AprA MT_L active sites
AprA MT_L-ΨGNAT 2Fo-Fc composite omit density contoured at 1σ. Maps were calculated with simulated annealing (starting temperature 5000 K). **a.** Metal-free MT_L-ΨGNAT, 2.4 Å. **b.** Mn^{2+} complex, 1.78 Å. **c.** Mn^{2+} and malonate complex, 1.85 Å.

domain (residues 1–227) is a large helical lid capping the MT_L core. As MTs typically include a lid domain, hereafter we refer to the AR-MT_L didomain as MT_L. The substantial sequence conservation within the lid consists of hydrophobic side chains that pack in the domain interior. A highly conserved 13-residue linker (residues 228–240) connects the MT_L lid and core, and is adjacent to the SAM binding site, but was disordered and lacked electron density.

The AprA GNAT (residues 505–621) is closely associated with MT_L, with only four amino acids between the last β-strand of the MT_L core and the first β-strand of the GNAT. The GNAT packs against a hydrophobic surface of MT_L helix 22, perhaps explaining why attempts to produce MT_L without GNAT resulted in insoluble protein. Interestingly, the AprA and CurA GNAT domains are highly similar for the first ~90 amino acids, after which the structures and sequences diverge abruptly (Figure 4.10). AprA GNAT is ~70 amino acids shorter than the CurA GNAT, specifically lacking the substrate binding tunnel and the His and Thr residues critical for decarboxylation (Figure 4.11a) [31]. Attempts to produce a longer MT_L-GNAT resulted in insoluble protein. Based on the structure superposition and sequence alignment, we conclude that the AprA GNAT is not a decarboxylase, and moreover, appears to be a remnant protein lacking a catalytic role and can be reclassified as a pseudo-GNAT (ΨGNAT).

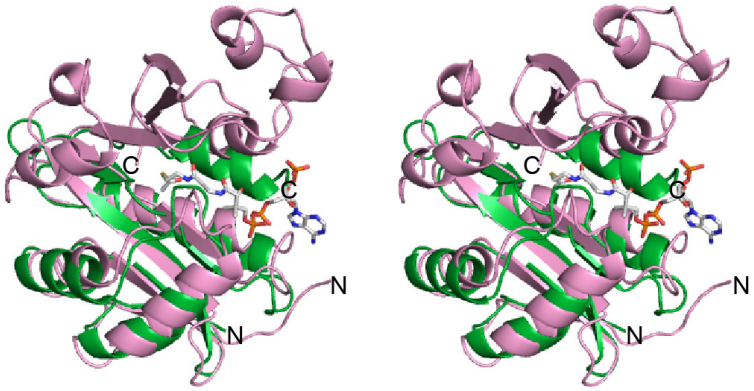


Figure 4.9 Superposition of AprA ΨGNAT and CurA GNAT
Stereo view of AprA ΨGNAT (green) superimposed with the CurA GNAT (pink) acetyl-CoA (sticks with white C) complex. The upper regions of CurA GNAT that are missing in AprA ΨGNAT include the lid of the CoA binding tunnel.

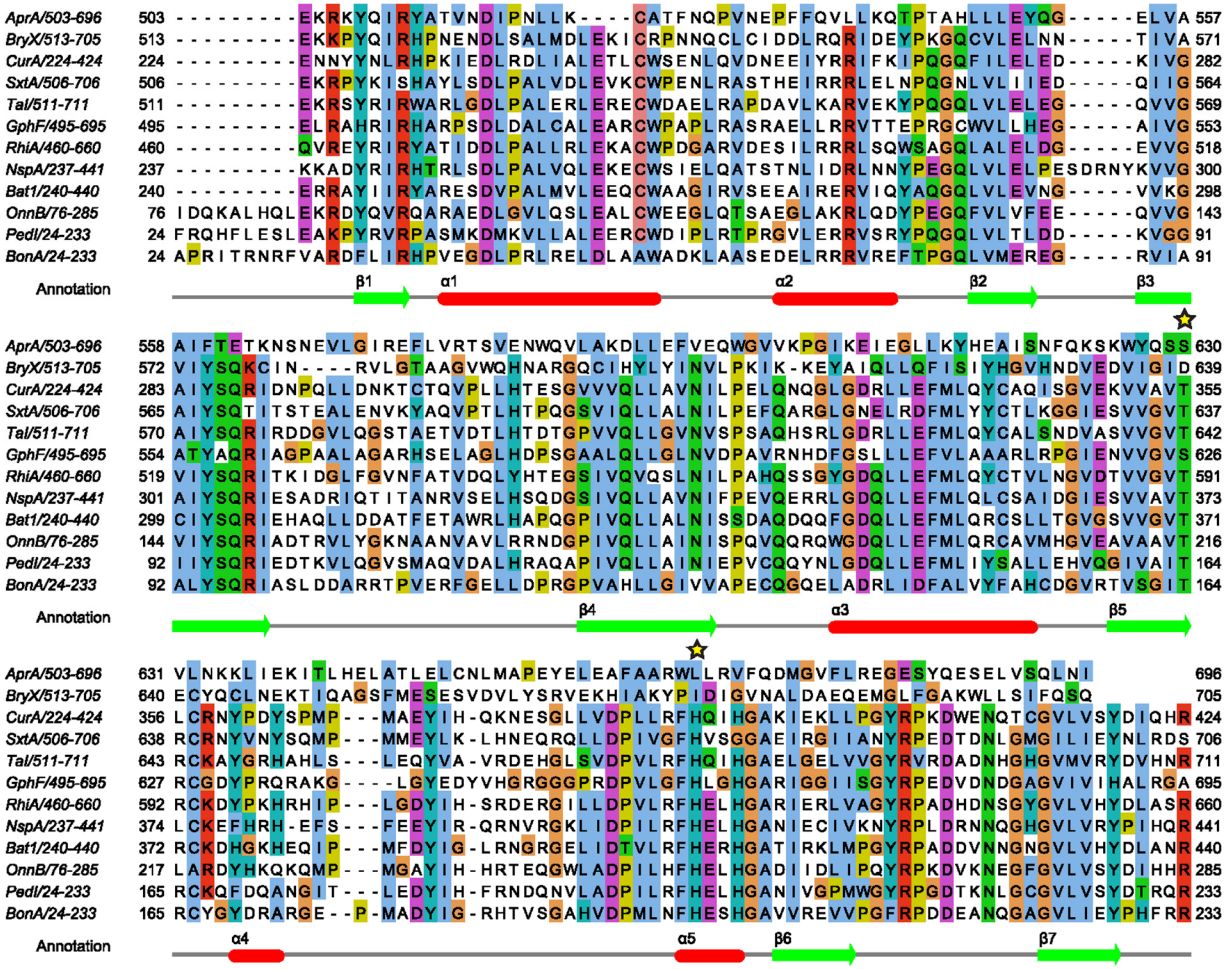


Figure 4.10 GNAT sequence alignment with CurA GNAT secondary structure annotation
Amino acids required for decarboxylation are starred. Pathway abbreviations (GenBank accession codes) are as follows: Apr- apratoxin A (WP_075900460), Bry- bryostatin (ABK51302.1), Cur- curacin A (AEE88289.1), Sxt- saxitoxin (WP_009343302.1), Ta- myxovirescin A (WP_011553948.1), Gph- gephyronic acid (KF479198.1), Rhi- rhizoxin (WP_013435483.1), Nsp- nosperin (ADA69237.1), Bat- batumin (WP_052451043.1), Onn- onnamide (AAV97870.1), Ped- pederin (AAR19304.1), Bon- bongkrekic acid(AFN27480.1)

Structural homology of MT$_L$

Among all entries in the structure database, the AprA MT$_L$ core domain structure is most similar to the two PKS *C*-methyltransferases, the CurJ MT domain (PDB 5THY, 13% sequence identity, RMSD 3.0 Å for 98 core Cα atoms) (Figure 4.11) and the fungal-derived citrinin MT domain (5MPT, 13% identity, RMSD 2.0 Å for 112 Cα atoms), which methylate the β-keto intermediate produced by a KS condensation reaction [117, 165]. Another homolog is MppJ (4KIG, 15% identity, RMSD 2.9 Å for 121 Cα atoms), a methyltransferase from the mannopeptimycin

biosynthetic pathway in *Streptomyces hygroscopicus*. MppJ uses a mononuclear iron, coordinated by two histidines and two waters bridged to an aspartic acid, to facilitate *C*-methylation at the benzylic position of phenylpyruvate [74]. In the AprA MT_L-ΨGNAT structure, His369 and His456 near the SAM binding site appeared to be positioned perfectly to serve as metal ligands similar to those in MppJ.

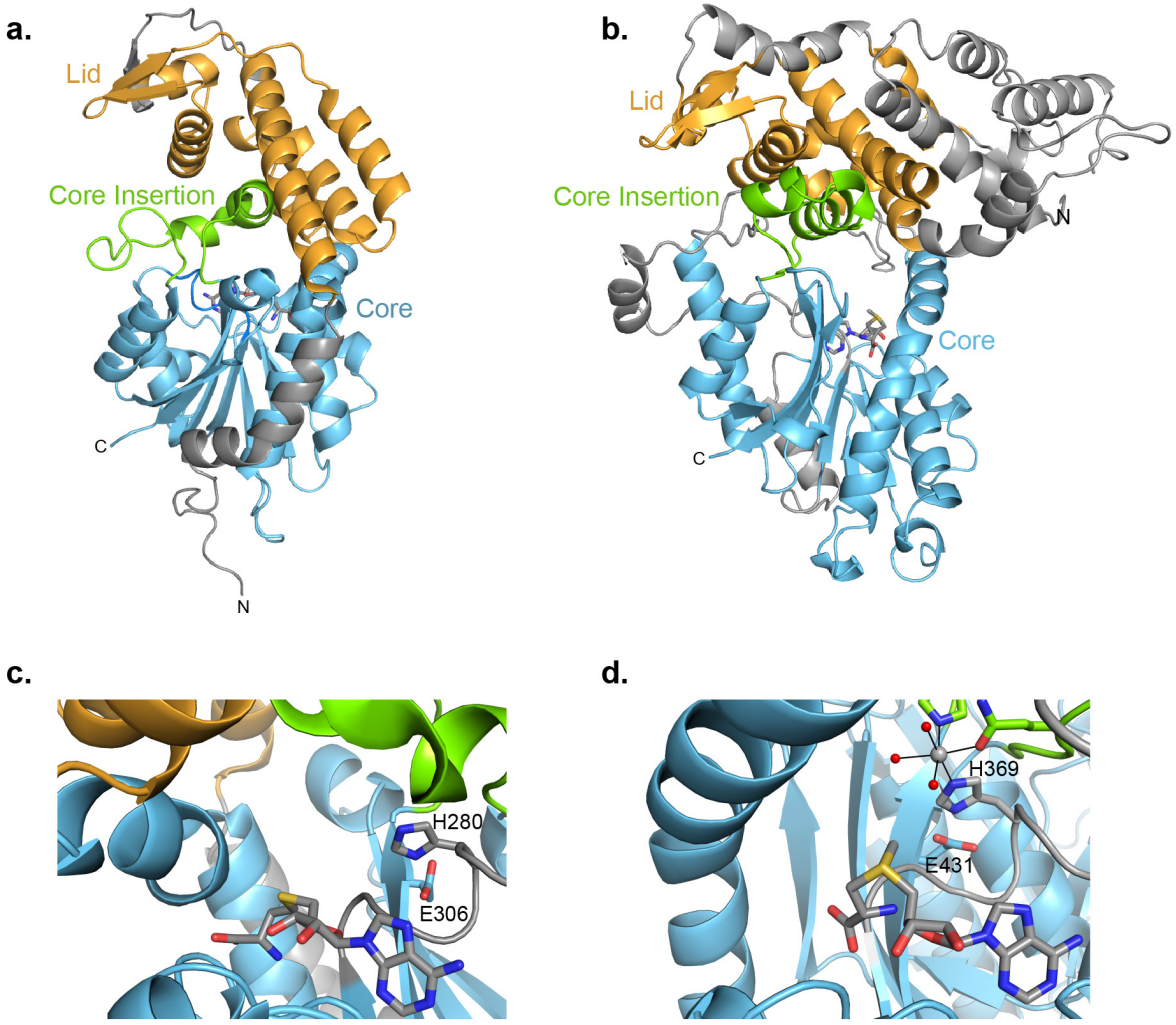


Figure 4.11 MT_L and *C*-MT homology
a. CurJ *C*-MT. Features in common with AprA MT1 are colored by region (lid orange; core blue; insertion green) and additional regions in gray. **b.** AprA MT1, colored as in b. **c.** CurJ *C*-MT active site. Protein colors as in b with SAH and the His-Glu catalytic dyad in sticks with gray C. **d.** AprA MT1 active site, colored as in d with Mn^{2+} (gray) and water (red) as spheres. His and Glu residues are in the same topological location as the CurJ *C*-MT His-Glu catalytic dyad.

AprA MT_L catalyzes iron-dependent methyltransfer

AprA MT_L-ΨGNAT was tested for SAM-dependent methyl transfer activity on a variety of acyl-ACPs (Mal, MeMal, acetyl, propionyl, isobutyryl) using excised AprA ACP. Reaction mixtures were monitored for product formation by mass spectrometry (MS) detection of ACP species. The phosphopantetheine (Ppant) ejection assay, which dissociates the Ppant fragment from the ACP phosphoserine linkage during ionization, was used for quantification based upon the relative abundance of Ppant species [150, 169]. No activity was observed for any acyl-ACP substrate tested. Inspired by the metal dependence of MppJ, we screened a panel of biologically relevant metals for their effect on activity (Figure 4.12, Figure 4.13, Figure 4.14). AprA MT_L -ΨGNAT catalyzed methylation of Mal-ACP to MeMal-ACP in the presence of Fe^{2+}, Mn^{2+}, Co^{2+}, and Ni^{2+}. With added Fe^{2+} under aerobic conditions, MT_L-ΨGNAT had a nearly three-fold higher activity than with the other active metals. Additionally aerobic reconstitution with Fe^{2+} yielded significant quantities of Me_2Mal-ACP (Figure 4.12b). As Fe^{2+} readily oxidizes to Fe^{3+} in aerobic environments and as activity is greatly reduced under anaerobic conditions, we infer that Fe^{3+} is the active metal.

AprA MT_L-ΨGNAT also methylated Mal- and MeMal-CoA, but had twofold lower activity with CoA-linked than with ACP-linked substrates under similar reaction conditions (Figure 4.15). Like AprA MT_L, GphF MT_L also catalyzed Mn- and SAM-dependent methylation of Mal-ACP, and produced both MeMal- and Me_2Mal-ACP with Fe^{3+} (Figure 4.16). This GphF MT_L activity is consistent with the predicted isobutyryl starter unit of gephyronic acid biosynthesis[21].

In contrast to the methylation of Mal-ACP, we detected no AprA MT_L methylation of acetyl, propionyl, or isobutyryl-ACP in the presence of Fe^{3+} (Figure 4.17a- c). There was also no detectable decarboxylation of Mal-, MeMal- or Me_2Mal-ACP in mass spectra of intact ACP species (Figure 4.13b), nor any enzyme-catalyzed acyl transfer from Mal-CoA or MeMal-CoA to the AprA ACP (Figure 4.17d, e), confirming that AprA ΨGNAT has no catalytic activity.

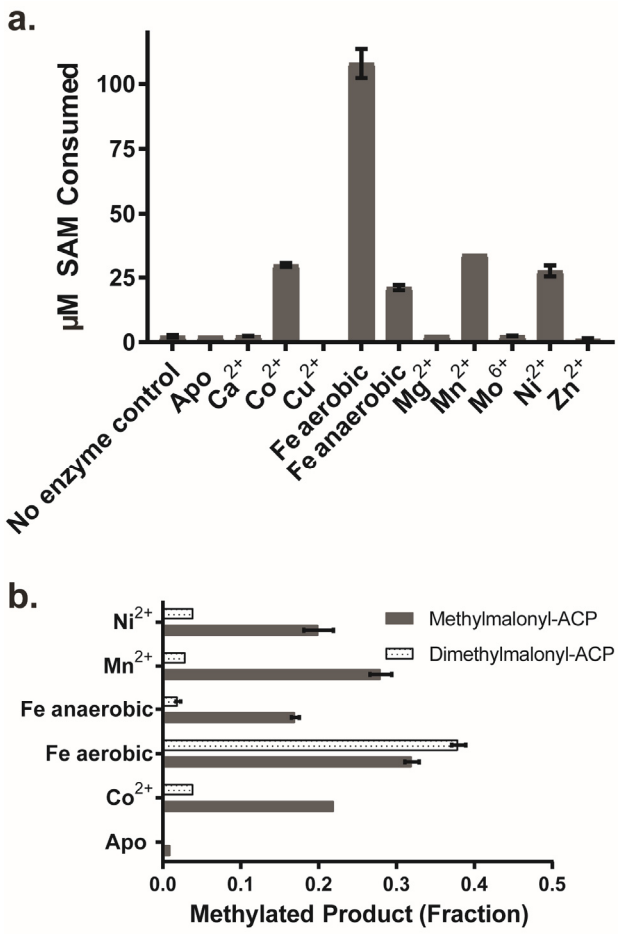


Figure 4.12 AprA MT_L-ΨGNAT activity.
a. Metal profile of reactions with Mal-ACP. The percent of Ppant ejection fragments from MeMal- and Me_2Mal-ACP are represented as equivalents of SAM consumed. **b.** Ratios of MeMal-ACP and Me_2Mal-ACP products of the reactions in a. Error bars represent triplicate experiments and, in some cases, are too small to be visible.

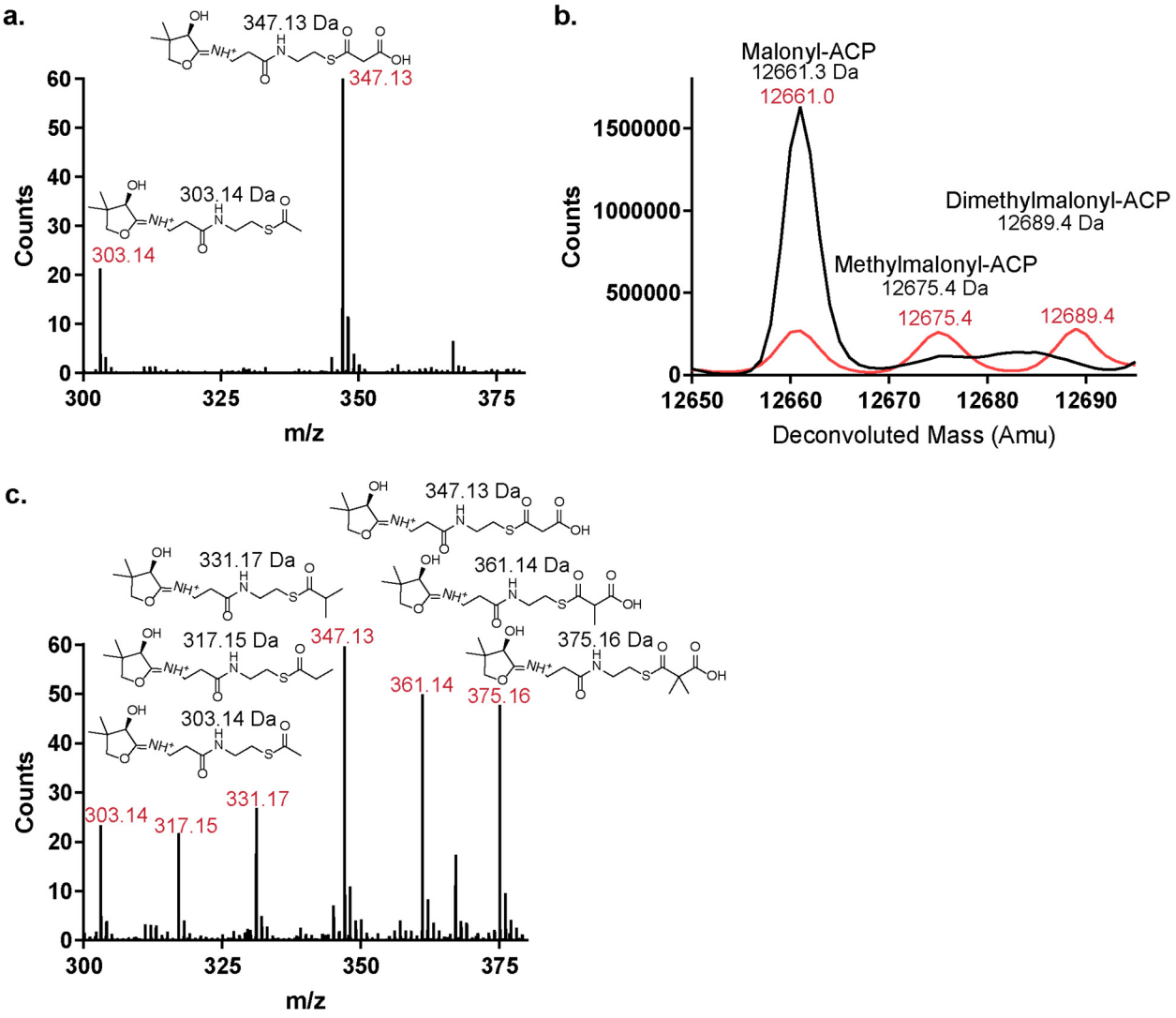


Figure 4.13 Representative electrospray-injection (ESI) mass spectra of AprA ACP from MT_L reaction mixes

Ppant ejection mass spectra and intact protein mass spectra are shown with experimental masses in red and calculated masses in black. **a.** AprA MT_L-ΨGNAT reaction with Mal-ACP and no added metal; Ppant ejection data. Malonyl fragments undergo in-source decay, producing the decarboxylation product acetyl. **b.** AprA MT_L-ΨGNAT reaction with Mal-ACP and Fe; intact ACP spectra. The no-metal control is in black; reaction with Fe in red. No masses were observed for acetyl-, propionyl-, or isobutyryl-ACP indicating 1) that AprA MT_L-ΨGNAT does not catalyze decarboxylation, and 2) that the decarboxylated products in the Ppant ejection spectra formed in the spectrometer not on the enzyme. **c.** AprA MT_L-ΨGNAT reaction with Mal-ACP and Fe; Ppant ejection data. The carboxylated fragments undergo in-source decay, producing acetyl, propionyl, and isobutyryl fragments.

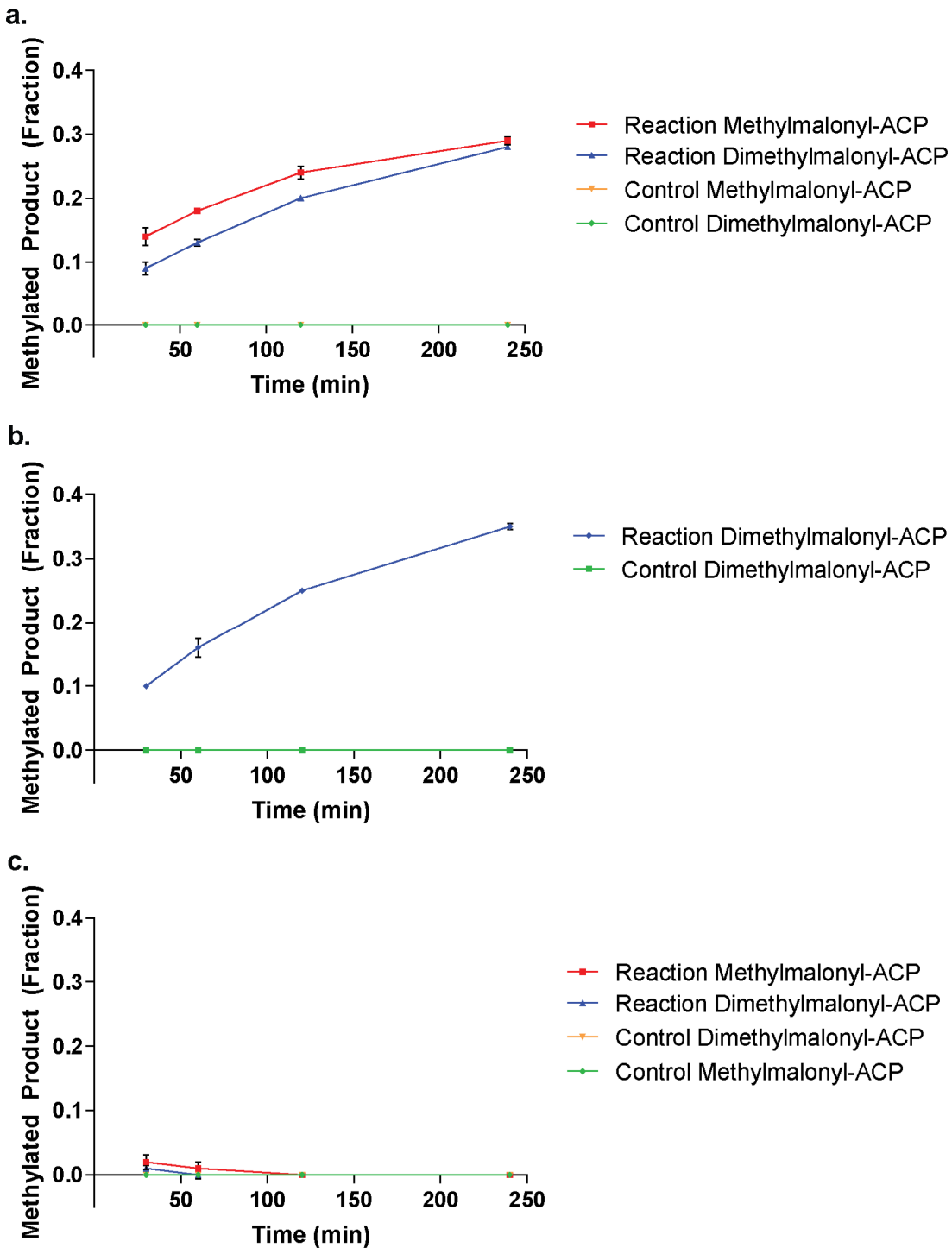
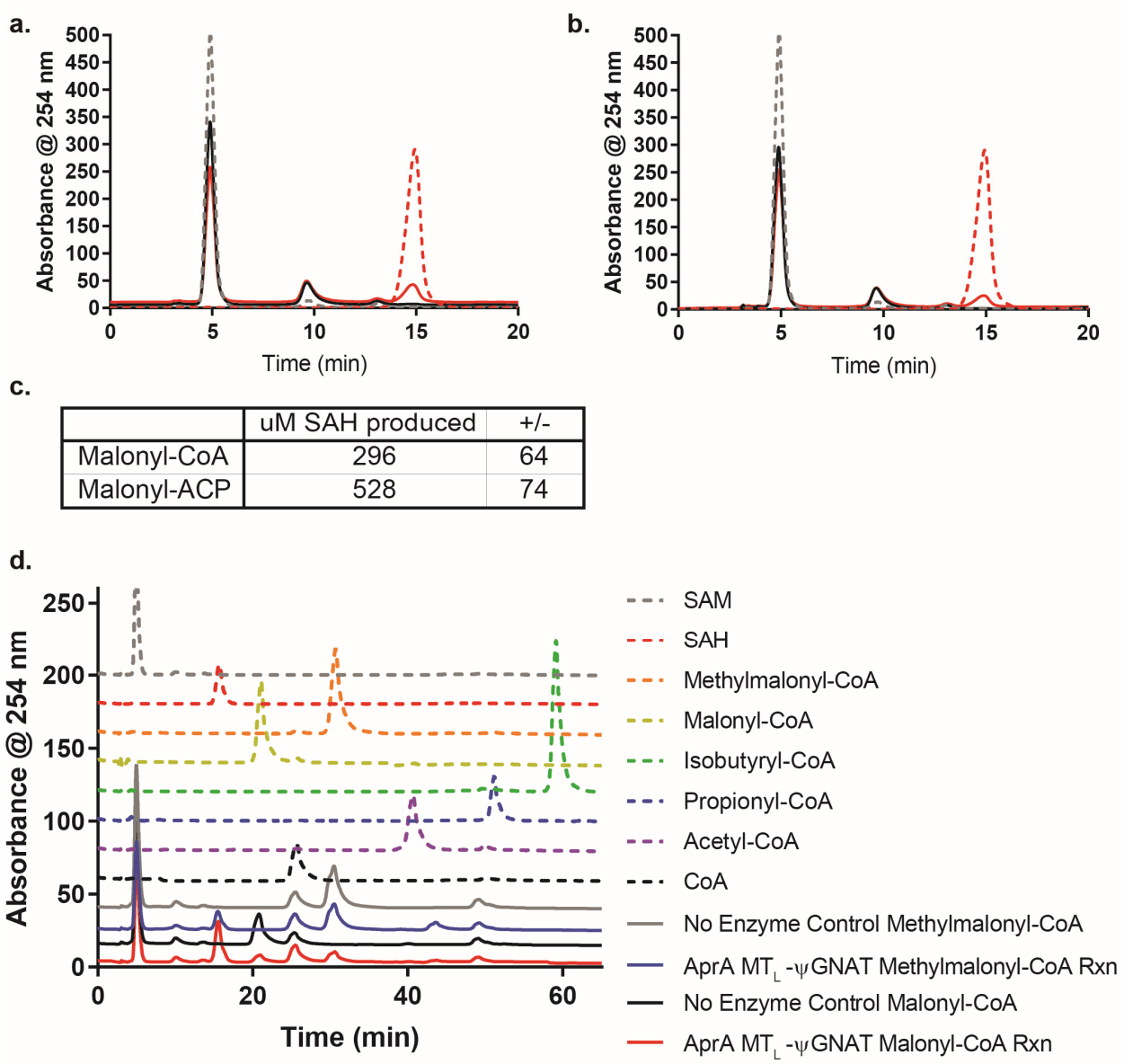


Figure 4.14 Time course data for AprA MT_L-ΨGNAT methylation reactions
a. AprA MT_L-ΨGNAT with Mal-ACP, **b.** AprA MT_L-ΨGNAT with MeMal-ACP, **c.** AprA ΨGNAT with Mal-ACP. Error bars represent triplicate experiments and, in some cases, are too small to be visible.



	uM SAH produced	+/-
Malonyl-CoA	296	64
Malonyl-ACP	528	74

Figure 4.15 AprA MT$_L$-ΨGNAT reactions with Mal-ACP, Mal-CoA or MeMal-CoA analyzed by HPLC

a. AprA MT$_L$-ΨGNAT with Mal-ACP substrate. Reaction, solid red; No enzyme control, solid black; SAM standard, dotted gray; SAH standard, dotted red. **b.** AprA MT$_L$-ΨGNAT with Mal-CoA substrate. Reaction, solid red; No enzyme control, solid black; SAM standard, dotted gray; SAH standard, dotted red. **c.** Calculated SAH produced in a and b based on absorbance at 254 nm. **d.** End-point HPLC analysis of Mal-CoA and MeMal-CoA reactions with AprA MT$_L$ -ΨGNAT. No enzyme controls and reactions, solid lines; standards, dotted lines.

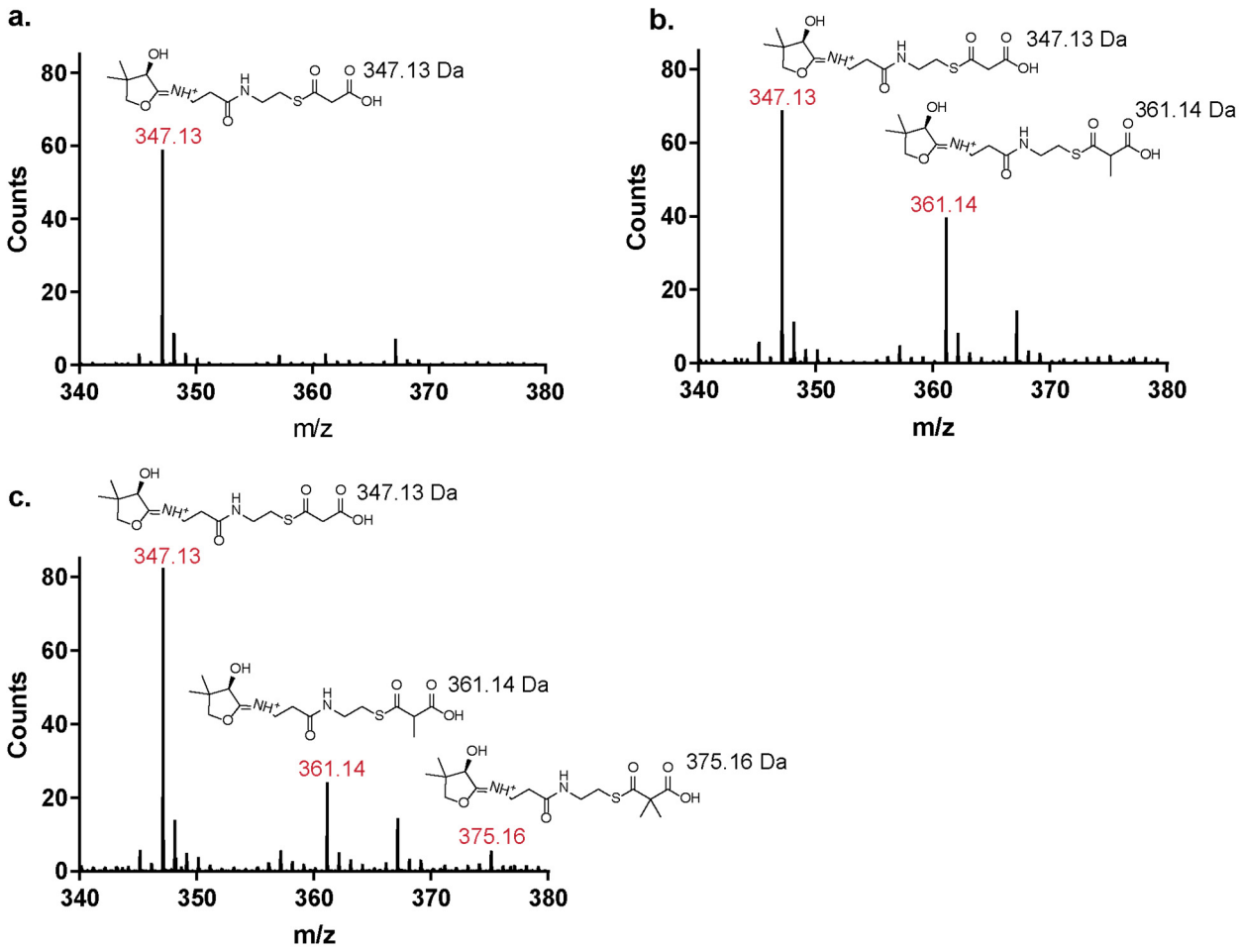


Figure 4.16 GphF MT_L-GNAT reaction with Mal-ACP

For the Ppant ejection data, experimental masses are in red and calculated masses in black. The GphF decarboxylase activity was blocked by using GphF MT_L-GNAT/H660A for these experiments. **a.** GphF MT_L-GNAT/H660A with Mal-ACP and no added metal. **b.** GphF MT_L-GNAT H660A with Mal-ACP and Mn. **c.** GphF MT_L-GNAT/H660A with Mal-ACP and Fe. The mass at 367 Da is a contaminant present in all samples.

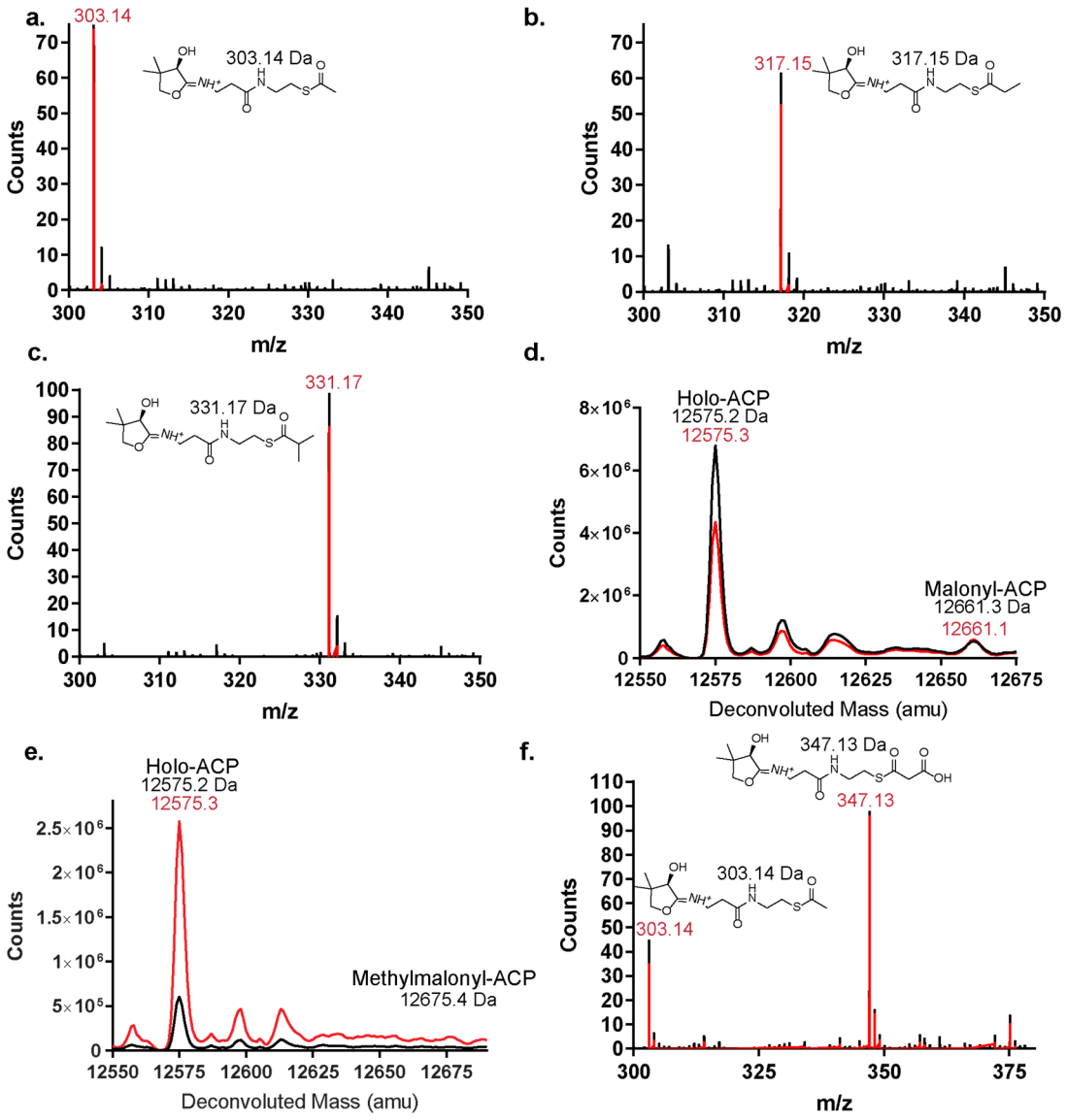


Figure 4.17 Mass spectra of methylation and acetyl transfer reactions with AprA MT$_L$-ΨGNAT, AprA ΨGNAT, and CurJ *C*-MT

Representative electrospray-injection (ESI) mass spectra of AprA ACP from MT$_L$-ΨGNAT and ΨGNAT reaction mixes and JamJ ACP from CurJ *C*-MT reaction mixes. Ppant ejection mass spectra are shown with experimental masses in red and calculated masses in black. **a.** AprA MT$_L$-ΨGNAT reaction with acetyl-ACP and Fe; Ppant ejection data. No-metal control, black; reaction, red. **b.** AprA MT$_L$-ΨGNAT reaction with propionyl-ACP and Fe; Ppant ejection data. No-metal control, black; reaction, red. **c.** AprA MT$_L$-ΨGNAT reaction with isobutyryl-ACP and Fe; Ppant ejection data. No-metal control, black; reaction, red. **d.** Intact protein spectra of AprA ΨGNAT acyltransfer reaction with holo-ACP and Mal-CoA. No-enzyme control, black; reaction, red. **e.** Intact protein spectra of AprA ΨGNAT acyltransfer reaction with holo-ACP and MeMal-CoA. No-enzyme control, black; reaction, red. **f.** CurJ *C*-MT reaction with Mal-ACP and Fe; Ppant ejection data. No-enzyme control, black; reaction, red.

Metal binding triggers tunnel formation

The metal in MT_L is labile; AprA MT_L purified without bound metal, and after reconstitution, did not retain iron during size exclusion chromatography. This property contrasts with MppJ, which co-purified with Fe^{3+} [74]. Metal addition blocked AprA MT_L-ΨGNAT crystallization in the previously identified conditions, but new conditions with 5 mM $MnCl_2$ resulted in a new crystal form and a 1.8-Å structure of AprA MT_L-ΨGNAT S274I/Q528P (Figure 4.6b and Table 4.2). Mn^{2+} is bound ~6 Å from the SAM methyl donor. The octahedral coordination sphere consists of three water ligands and three protein ligands (Figure 4.7b, Figure 4.18), including Gln461 and the predicted His369 and His456. These three ligands are conserved in MT_L homologs (Figure 4.5), the only exceptions being a few cases of Gln in place of His456. No charged ligand resides in the first coordination shell, but one water ligand is bridged to Asp370. A D370N substitution resulted in a slight decrease in activity (Figure 4.19).

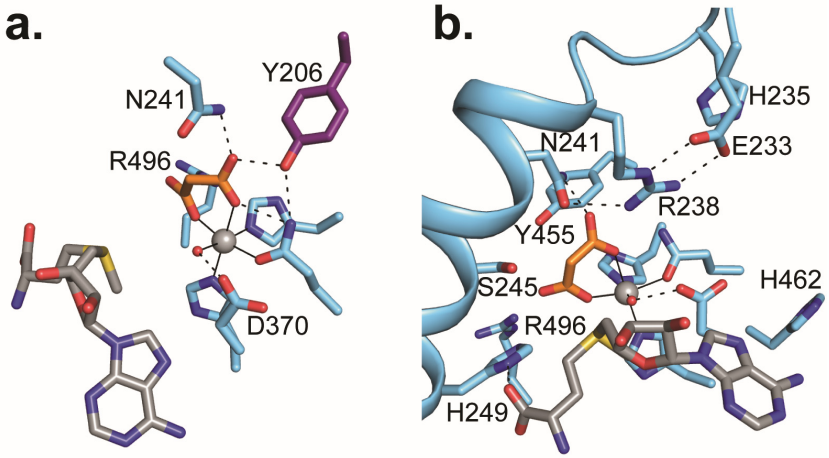


Figure 4.18 Key amino acids for substrate binding
a. Malonate interactions at the metal center. **b.** Hydrogen bond network in lid-core connector (amino acids 228–251). Amino acid side chains (purple C, blue C), SAM (gray C), and malonate (orange C) are shown in sticks with atomic coloring, hydrogen bonds as dashed lines, and metal coordination bonds as solid lines.

Metal binding induced a substantial reorganization of the active site, especially in the conserved linker between the MT_L lid and core (residues 228–240), which was disordered in the metal-free structure. Amino acids 228–251 form a helix (helix 14) and loop that extend over the metal center and SAM binding site, and create a tunnel between the lid and core (Figure 4.6b, Figure 4.7b). Surprisingly, intact SAM was bound in the Mn^{2+} complex, suggesting that the fully

formed active site protects SAM from hydrolysis. In contrast to the metal-free structure, the methionine portion of SAM was well ordered (Figure 4.7b, Figure 4.8b). The SAM carboxylate is hydrogen bonded with the His249 imidazole, which points into the active site between SAM and Mn^{2+} (Figure 4.18). A H249A substitution abolished activity with Mal-ACP substrates (Figure 4.19). Second-shell interactions at the metal center are also important to catalysis. The Gln461 and His369 metal ligands are positioned by conserved residues Tyr206 in the lid domain, and Glu431 in the core. A Y206F substitution eliminated activity (Figure 4.19). His462 lies on a loop following His456 and Gln461 and appears to stabilize these ligands through hydrogen bonds with adjacent loops. Substitution at His462 resulted in a twofold decrease in activity (Figure 4.19).

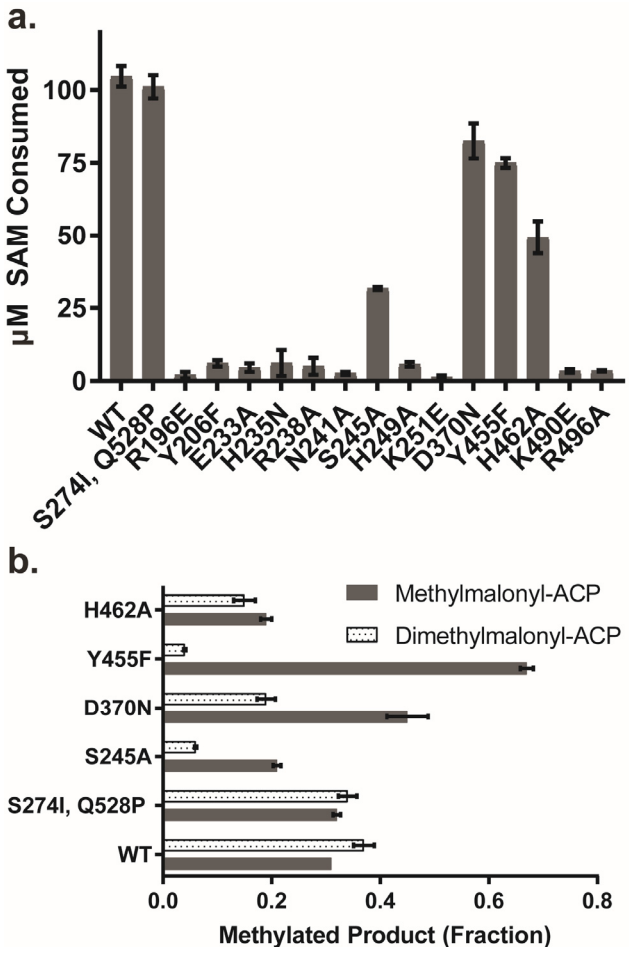


Figure 4.19 Relative methylation activities of wild type AprA MT_L-ΨGNAT, serendipitous crystallization substitution (S274I/Q528P), and active site variants
a. SAM consumption in reactions with Mal-ACP. **b.** Ratios of MeMal-ACP and Me_2Mal-ACP products from a. Error bars represent triplicate experiments and, in some cases, are too small to be visible.

Substrate binding and catalysis

In order to understand how the metal center promotes catalysis, we obtained a 1.85-Å structure for an MT_L-ΨGNAT complex with Mn^{2+}, SAM and malonate (Figure 4.7c, Figure 4.18, Table 4.2). Malonate binds directly to Mn^{2+} as a bidentate ligand in which one coordination bond from each carboxylate displaces a water ligand. One malonate carboxylate is clearly in the position of the terminal group of Mal-ACP as it forms three hydrogen bonds with conserved amino acids: Tyr206, Asn241 and Gln461, which are in turn positioned by an extensive hydrogen bond network with other conserved residues [Ser205 in the lid, and Glu233 and Arg238 in the re-ordered lid-core connector, (Figure 4.18)]. Methyl transfer activity was abolished by substitution of Asn241 or any connector amino acids in the hydrogen bond network, including His235 (Figure 4.19). The second malonate carboxylate is well positioned to form the thioester with the ACP Ppant as it lacks hydrogen bonds to the protein and the non-coordinated oxygen points into the tunnel between the lid and core. Like the metal center, the malonate has no direct contact with charged amino acids. This should enhance the ionic interaction of malonate and metal, and the ability of the metal to promote methyl transfer.

Metal-dependent MTs typically require Ca^{2+} or Mg^{2+}, which functions to position the substrate for methyl addition and proton abstraction from a relatively acidic oxygen [65, 67, 189]. Compared to a hydroxy oxygen, the weak acidity of the α-carbon renders Mal-ACP a far more challenging methylation target. The mannopeptimycin methyltransferase MppJ, which methylates phenylpyruvate at a benzylic carbon of similarly weak acidity [74], provides a direct precedent for the Fe^{3+}-dependence of the AprA MT_L.

We propose that the metal facilitates catalysis by acting as a Lewis acid to lower the pKa of the malonyl or methylmalonyl α-carbon, leading to deprotonation, enolate formation and attack of the SAM methyl (Figure 4.20), similar to the mechanism proposed for Co^{2+}-dependent epimerization of MeMal-CoA [190]. The positions of malonate and SAM indicate that the product of the first methyl transfer is (*S*)-MeMal-ACP (Figure 4.7c, Figure 4.18). The malonyl α-carbon is 6.8 Å from the SAM methyl, suggesting that further closure of the MT_L active site occurs prior to methyl transfer. Tyr455 and Ser245 are the only amino acids in the active site positioned (5.0 Å and 4.1 Å, respectively) to accept an α-proton. Interestingly, a Y455F substitution and an S245A substitution abrogated the second methyl transfer but not the first (Figure 4.19b). This raises the possibility that one of these side chains accepts the α(*R*)-proton in the second methyl transfer

reaction and that another base accepts the α(*S*)-proton prior to the first methyl transfer. The only candidate for an α(*S*)-proton acceptor is the water molecule that bridges Asp370 and the metal (3.5 Å from the malonyl α-carbon).

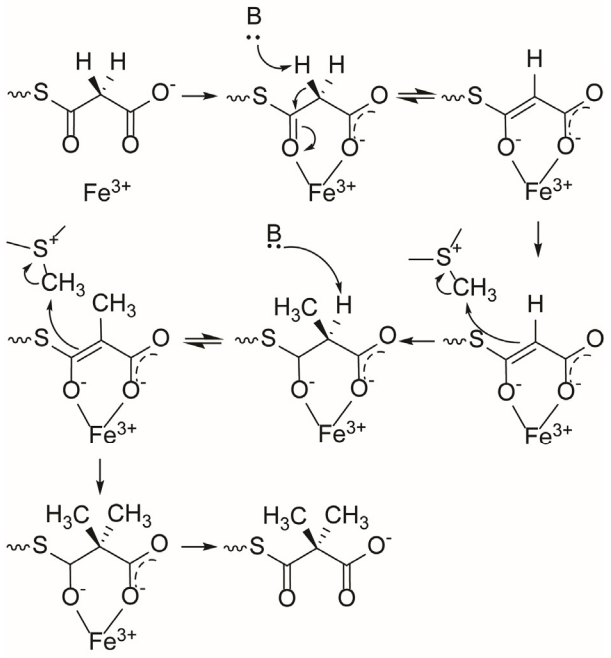


Figure 4.20 AprA MT_L mechanism
Fe^{3+} acts as a Lewis acid to promote methylation at the α-carbon of Mal-ACP to produce (*S*)-MeMal-CoA. MeMal-ACP is subsequently methylated to Me_2Mal-ACP.

The few examples of mononuclear iron as a Lewis acid include nitrile hydratase and the MppJ MT, which both employ Fe^{3+} [74, 191, 192]. Ferric iron supports AprA MT_L catalysis, as our assays were performed in an aerobic environment over several hours and activity was diminished under anaerobic conditions (Figure 4.12). Although AprA can perform monomethylation with Co^{2+}, Fe^{2+}, Mn^{2+}, and Ni^{2+}, dimethylation is dependent on Fe^{3+} as a co-factor (Figure 4.12b). We infer that the second AprA MT_L methylation is more energetically demanding than the first due to the difference in pKa values reported for methyl ester forms of malonate (15.9) and methylmalonate (18.0) [193]. Ferric iron may be uniquely capable of facilitating the removal of the second proton for the formation of Me_2Mal-ACP, as it is a stronger Lewis acid than Co^{2+}, Fe^{2+}, Mn^{2+} and Ni^{2+}. Despite the biosynthetic challenge, nature has evolved other approaches to generate the dimethylmalonyl biosynthetic subunit, as two non-metal dependent PKS extension module *C*-MTs have also been reported to produce Me_2Mal-ACP [100, 101, 194].

Ppant modeling

We modeled Mal-Ppant into the AprA MT_L structure based on the outermost carboxylate of bound malonate (Figure 4.21a). The tunnel width and length are well matched for threading Ppant to deliver malonyl to the active site. The Ppant could interact with conserved tunnel residues Gln248 and Arg496, which coordinates a glycerol molecule from the crystallization solution. Although Arg496 is 10 Å from the metal center, an Ala substitution abolished activity (Figure 4.19a), suggesting that Ppant interactions within the tunnel are essential for activity.

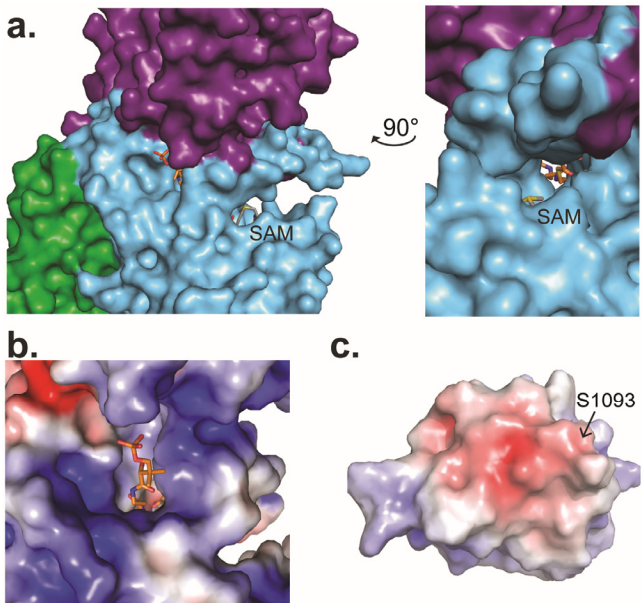


Figure 4.21 Malonyl-Ppant substrate modeling
a. MT_L-ΨGNAT surface showing Ppant, SAM and the tunnel created by active site assembly. Ppant (orange C) and SAM (gray C) are shown in sticks. Modeled Mal-Ppant spans tunnel between the MT_L lid (purple) and core (blue). SAH can exchange with SAM while the substrate is still bound in the active site. **b.** Electrostatic surface of AprA **c.** Electrostatic surface of homology model of AprA GNAT. Ser1093 is the Ppant attachment site. Electrostatic surfaces shown at 5 kT/e, blue electropositive, red electronegative.

The proposed Ppant tunnel entrance is surrounded by positively charged amino acids (Arg196, Lys251, Lys290). R196E, K251E and K490E variants yielded no activity (Figure 4.19a). A complementary negatively charged surface is adjacent to the Ppant attachment site (Ser1093) in a homology model for AprA ACP. This charge complementarity may facilitate the interaction between AprA MT_L and ACP (Figure 4.21b, c). Ppant binding in the tunnel is compatible with exchange of SAH for SAM through the opposite end of the tunnel, facilitating the second methylation reaction prior to disengagement of MeMal-ACP (Figure 4.21a). Evidence that two

methylations occur prior to dissociation of ACP from the enzyme comes from the nearly identical accumulation rates of MeMal-ACP and Me_2Mal-ACP in a reaction starting with Mal-ACP, indicating no initial accumulation of the singly methylated product, MeMal-ACP (Figure 4.14a).

Relation of AprA MT_L to other PKS MTs

Of the several hundred MT structures in the structure database, AprA MT_L most closely resembles the MTs from PKS extension steps [117, 165]. All class I SAM-dependent MTs are thought to have a common ancestor, but the superfamily has several highly diverged branches. AprA shares several features with the PKS extension module MTs (Figure 4.11a, b) [117], including a lid-interacting insertion between β-strands five and six of the core. The lid domains are topologically similar at the N-termini, but the AprA MT_L lid (formerly known as AR) is much larger due to a C-terminal extension. We previously showed that conserved His and Glu amino acids are essential to methyl transfer by the CurJ extension-module MT, and proposed that the His imidazole is the catalytic base that accepts a proton from the α-carbon of the β-ketoacyl-ACP substrate [117]. By structure superposition, the critical His and Glu of extension-module MTs correspond to AprA MT_L His369 and Glu431, which are also conserved in GNAT initiation modules (Figure 4.11c, d). His369 is a metal ligand, and Glu431 is hydrogen bonded to the second imidazole nitrogen of His369, exactly as in the CurJ MT. Interestingly, CurJ MT acts on acetoacetyl-ACP, but has no activity with Mal-ACP, even with the addition of iron (Figure 4.17f) and lacks the other metal ligands (His456 and Gln461). Thus, it appears that an ancestral extension-step MT acquired a metal center as it adapted from α-methylation of β-ketoacyl-ACP to the more energetically demanding α-methylation of β-carboxyacyl-ACP.

The dual entrances to the active site are a major difference between AprA MT_L and both the PKS extension *C*-MTs, which have a single active-site entrance. In extension module *C*-MTs, the Ppant tunnel entrance is blocked by an N-terminal helix [117]. Thus, it appears that AprA MT_L has specifically evolved two active site entrances, which may be essential to performing the dimethylation reaction.

Functional annotation of GNAT initiation modules

The discovery of iron-dependent methylation of Mal-ACP by AprA MT_L and homologs permits the functional annotation of branched-chain acyl group production by GNAT initiation modules for several PKS pathways, including bryostatin (pivalate), gephyronic acid (isobutyryl),

myxovirescin (propionyl) and saxitoxin (propionyl). The initiation modules that generate pivaloyl-ACP, AprA and BryX, have an MT_L-ΨGNAT-$MT2_L$-ACP architecture, as the ΨGNAT lacks the His and Thr/Ser residues required for decarboxylation and also the substrate tunnel for Mal-CoA binding [31]. Consistent with the structure and sequence we observed no decarboxylation of Mal-, MeMal- or Me_2Mal-ACP in assays of AprA MT_L-ΨGNAT. Additionally, no Mal- or MeMal-CoA transfer to the ACP was observed by the ΨGNAT (Figure 4.17d, e). Thus, for the MT_L-ΨGNAT-$MT2_L$-ACP initiation modules AprA and BryX, the identity of the enzyme responsible for activating and catalyzing transfer of the malonyl group onto the ACP is unknown, as is the identity of the corresponding decarboxylase.

In modules containing MT_L-GNAT-ACP, such as GphF, TaI and SxtA (Figure 4.1b, c), MT_L catalyzes one or two methyl transfer reactions on Mal-ACP. No evident sequence motifs or structural features distinguish monomethylating and dimethylating MT_L enzymes. The MeMal- or Me_2Mal-ACP product should be decarboxylated by GNAT to yield propionyl- or isobutyryl-ACP starter units, as all GNAT domains in MT_L-GNAT-ACP modules contain the His and Thr/Ser side chains required for decarboxylation [31]. Unlike the CurA GNAT, the GNAT of pathways with branched-chain starter units should not catalyze decarboxylation of Mal-CoA. These GNATs may transfer malonyl from CoA to ACP to initiate starter unit biosynthesis, however this would require the ACP-bound substrate to occupy the GNAT active site both before and after MT_L methylation. This process could additionally lead to aberrant product formation through premature decarboxylation of Mal-ACP. Therefore for propionyl- or isobutyryl-ACP production, we reason that an enzyme outside the module transfers malonyl to the ACP. For pathways with *in trans* acyltransferases (AT), one or more loading enzymes are encoded in the gene cluster. For pathways with *in cis* ATs, the loading enzyme is unknown, but may be a malonyl acyltransferase from fatty acid biosynthesis, as observed in other polyketide pathways [195-197].

Evolution of GNAT initiation modules

The extant GNAT initiation modules are remarkable examples of PKS evolution in action. AR- MT_L-GNAT-ACP modules are capable of synthesizing propionyl or isobutyryl starter units. In order to form a pivaloyl starter unit, a second methyltransferase ($MT2_L$), which is most similar to PKS extension-step *C*-MTs, is inserted into the module. However, it appears that the decarboxylation and acyltransfer functions of GNAT were lost in pivaloyl-producing modules, yielding a ΨGNAT. GNAT initiation modules that introduce an acetyl starter unit are presumed

to function like the CurA GNAT by decarboxylation of Mal-CoA and acetyl transfer to the ACP. Several of these initiation modules retain vestiges of an MT ancestor. For example, the MT_L core domain was lost from the curacin A initiation module (Figure 4.1f) but the AR, which we now know is a remarkably large MT_L lid, was retained and enhances, but is not required for, GNAT loading activity[31]. In contrast, GNAT initiation modules from the batumin (Bat) and nosperin (Nsp) pathways have lost the AR and contain only $MT_{L\ core}$-GNAT-ACP (Figure 4.1e) [93, 116]. The rhizoxin GNAT initiation module has homologs of all domains, but the MT_L appears to be non-functional (Figure 4.1d). Finally, minimal GNAT initiation modules in the bongkrekic acid (Bon), onnamide A (Onn), and pederin (Ped) pathways contain only GNAT and ACP (Figure 4.1g) [94-96].

In conclusion, we have characterized an unusual Fe^{3+}- and SAM-dependent methyltransferase involved in the production of branched polyketide starter units. Although related to other PKS MTs, the MTs of GNAT initiation modules have evolved to have two active site entrances to facilitate dimethylation. A crystal structure in complex with malonate, a substrate mimic, reveals an elegant conserved hydrogen bond network responsible for positioning the substrate. Functional and structural characterization of the AprA MT_L provides key insights to enable a more robust annotation of GNAT initiation modules found in diverse microorganisms.

Chapter 5 Biosynthesis of *t*-Butyl in Apratoxin A: Functional Analysis and Architecture of a PKS Loading Module

Reproduced with permission from Skiba, M.A., Sikkema, A.P., Moss, N.A., Lowell, A.N., Su, M., Sturgis, R.M., Gerwick, L., Gerwick, W.H., Sherman, D.H., Smith, J.L. Biosynthesis of *t*-Butyl in Apratoxin A: Functional Analysis and Architecture of a PKS Loading Module. *ACS Chem Biol* **13**, 1640-1650 (2018).

Summary

The unusual feature of a *t*-butyl group is found in several marine-derived natural products including apratoxin A, a Sec61 inhibitor produced by the cyanobacterium *Moorea bouillonii* PNG 5-198. Here we determine that the apratoxin A *t*-butyl group is formed as pivaloyl acyl carrier protein (ACP) by AprA, the polyketide synthase (PKS) initiation module of the apratoxin A biosynthetic pathway. AprA contains an inactive "pseudo" GCN5-related N-acetyltransferase domain (ΨGNAT) flanked by two methyltransferase domains (MT_L and $MT2_L$) that differ distinctly in sequence. Structural, biochemical, and precursor incorporation studies reveal that $MT2_L$ catalyzes unusually coupled decarboxylation and methylation reactions to transform dimethylmalonyl-ACP, the product of MT_L, to pivaloyl-ACP. Further, pivaloyl-ACP synthesis is primed by the fatty acid synthase malonyl acyltransferase (FabD), which compensates for the ΨGNAT and provides the initial acyl-transfer step to form AprA malonyl-ACP. Additionally, images of AprA from negative stain electron microscopy reveal multiple conformations that may facilitate the individual catalytic steps of the multienzyme module

Introduction

Marine organisms are rich sources of bioactive natural products[198, 199], providing potential leads for new pharmaceuticals. Cyanobacteria produce a myriad of polyketide and non-ribosomal peptide secondary metabolites, which are synthesized by polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) assembly lines using acyl-coenzyme A (CoA) or amino acid building blocks, respectively. Interestingly, a number of marine natural products contain *t*-butyl

groups[200-216], a chemical moiety that is relatively rare in nature[217]. Typically, *t*-butyl groups in natural products are encountered either as modified amino acids[201-204], terpene-derived pendant chains on the D-ring of steroids[205], or in PKS/NRPS-derived molecules[200, 206-212] primarily found in the metabolomes of sponges or cyanobacteria.

Although *t*-butyl-containing natural products were first identified over 50 years ago[218], only one route for *t*-butyl biosynthesis has been characterized[202, 203], which involves a cobalamin-dependent radical *S*-adenosylmethionine (SAM) enzyme. Apratoxin A[25] is a *t*-butyl containing cytotoxic Sec61 inhibitor (Figure 5.1a) produced by the marine cyanobacterium *Moorea bouillonii* PNG5-198[25-27]. However, the gene cluster for apratoxin A biosynthesis encodes no homolog of the cobalamin-dependent radical SAM enzymes previously implicated in *t*-butyl synthesis, which should occur in the initial steps of the pathway. Instead, synthesis of the *t*-butyl group in the form of pivalate is proposed to be carried out by AprA, an unusual polyketide initiation module containing a GCN5-related *N*-acetyltransferase (GNAT)-like domain flanked by two methyltransferase domains (MT_L and $MT2_L$; Figure 5.1b)[25]. GNAT-like domains typically decarboxylate CoA- or ACP-linked substrates[31, 183, 219], and the founding member was also associated with the subsequent acyl transfer from CoA to an ACP domain during the initiation of curacin A biosynthesis[31]. However, we previously showed that the AprA GNAT is truncated and lacks catalytic residues essential for decarboxylation, and thus re-annotated it as a "pseudo" GNAT (ΨGNAT) [119]. AprA MT_L and $MT2_L$ share very low amino acid sequence identity, but $MT2_L$ is more than 30% identical to *C*-methyltransferases (*C*-MT) found in some PKS extension modules. PKS *C*-MTs of this type, for example the CurJ *C*-MT, methylate the α-position of β-keto intermediates during cycles of polyketide chain extension and modification[117]. Several variants of the marine natural product bryostatin also have a *t*-butyl substituent[208], and a homolog of AprA (MT_L -ΨGNAT- $MT2_L$ -ACP) exists within BryX in the bryostatin pathway[20].

Previously, we characterized two initial steps in the biosynthesis of the apratoxin A starter unit, demonstrating that AprA MT_L is a mononuclear iron-dependent methyltransferase that forms dimethylmalonyl-ACP (Me_2Mal-ACP) from malonyl-ACP (Mal-ACP) and two equivalents of SAM (Figure 5.1c)[119]. Thus, conversion of Me_2Mal-ACP to the pivaloyl starter unit requires a decarboxylase and a methyltransferase. The acyltransferase that initiates apratoxin A biosynthesis by forming Mal-ACP (Figure 5.1c) also remains to be identified.

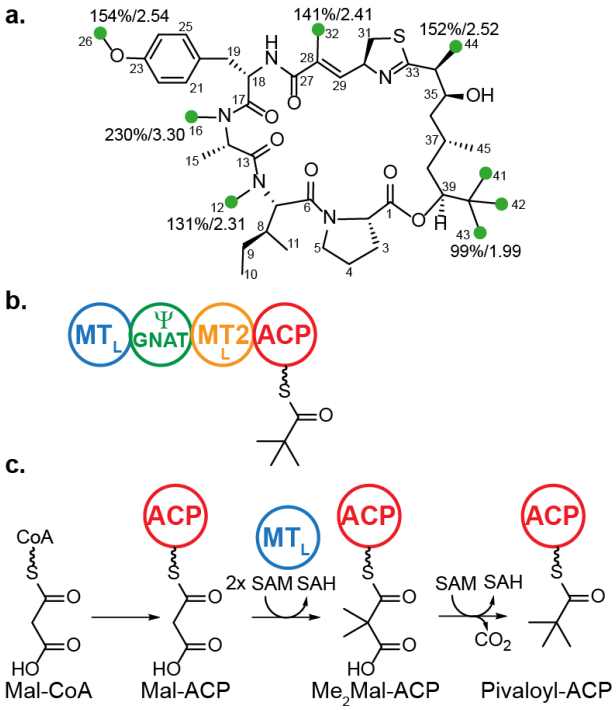


Figure 5.1 Production of a *t*-butyl group by AprA
a. Apratoxin A structure and distribution of SAM-derived methyl groups. Feeding studies demonstrate that the *t*-butyl group in apratoxin A is derived from SAM. Percent enrichment over natural abundance and fold-change of SAM-derived carbons (green circles) from [methyl-^{13}C]methionine-fed cultures of *M. bouillonii* PNG5-198 relative to native abundance apratoxin A are displayed. **b.** Cartoon representing AprA domains, which are proposed to produce pivaloyl-ACP: methyltransferase 1 (MT_L), pseudo GCN5-related *N*-acetyltransferase (ΨGNAT), methyltransferase 2 ($MT2_L$), ACP. **c.** Reactions needed to produce pivaloyl-ACP. The conversion of malonyl-ACP to Me_2Mal-ACP has been characterized[119]. The identities of the enzymes that catalyze acyltransfer from Mal-CoA to ACP, decarboxylation of Me_2Mal-ACP, and the third methylation reaction are addressed in this study.

To determine whether AprA can form pivaloyl-ACP, we characterized the catalytic activity of the $MT2_L$, solved a crystal structure of the ΨGNAT- $MT2_L$ didomain, and conducted stable-isotope labeled feeding experiments with live cultures of *M. bouillonii* PNG5-198. $MT2_L$ is a remarkable bifunctional enzyme that catalyzes the coordinated decarboxylation and methylation of Me_2Mal-ACP to produce pivaloyl-ACP. As no AprA catalytic domain can perform the missing malonyl-acyltransfer step to initiate the pathway, we determined that the *M. bouillonii* fatty acid synthase malonyl-acyltransferase, FabD, is an efficient catalyst for the formation of Mal-ACP from Mal-CoA, indicating crosstalk between primary and secondary metabolic pathways in *M. bouillonii.* Furthermore, a model of the AprA full module based on the crystal structures of the

AprA MT_L-ΨGNAT[119] and ΨGNAT- $MT2_L$ didomains was validated by negative-stain electron microscopy (EM) of AprA ΔACP in solution, revealing a flexible overall architecture that may facilitate the individual catalytic steps.

Experimental Procedures

Culturing, extraction, and purification of apratoxin A

All cyanobacterial culturing and extraction and purification of apratoxin A was performed by Nathan Moss. Culture conditions of *Moorea bouillonii* PNG 5-198 were identical for [methyl-^{13}C]methionine and labeled [1-^{13}C]propionate feeding experiments. Apratoxin A producer *M. bouillonii* PNG 5-198 was cultured in SW-BG11 media between 27-28°C, under light of 5.4-10.8 μmol photons m^{-2} S^{-1}. Two batches of 0.5 g wet filaments 1 cm in length were inoculated into 250 mL Erlenmeyer flasks and grown for 14 days, followed by sub-culturing of each biomass flask into two 2.8 L Fernbach flasks containing 1 L of media. After growth for ten days, the biomass from each flask was combined into one 2.8 L Fernbach flask containing 750 mL media, and the culture was equilibrated for an additional 7 days. [Methyl-^{13}C]methionine (Cambridge Isotopes) was added to the culture in three batches of 20 mg each over 13 days, to a final concentration of 0.53 mM, and harvested after an additional 6 days of growth. [1-^{13}C]propionate (Cambridge Isotopes) was added to a separate culture flask with a similar quantity of biomass in three batches of 28.8 mg over period of 15 days, to a final concentration of 1.2 mM, and harvested after an additional 15 days of growth. Cultures were harvested and lightly washed with DI water over a Büchner funnel with filter paper, followed by extraction with 2:1 CH_2Cl_2:MeOH. Crude extracts were fractionated by C18 SPE (Agilent) and eluted with 40%, 80%, and 100% acetonitrile, then 100% CH_2Cl_2. The 80% and 100% acetonitrile fractions were combined and further purified by HPLC using a gradient of 65% to 99% acetonitrile in H_2O + 0.1% formic acid at 3 mL min^{-1} through a 150 x 10 mm, 5 μM Kinetex C18 semi-preparative column. Subsequent semipure apratoxin A plus demethylated analogs were re-submitted for HPLC using a gradient of 70% to 81% acetonitrile in H_2O + 0.1% formic acid at 4 mL min^{-1}, rendering 1.2 mg of apratoxin A from the [1-^{13}C]propionate experiment and 1.8 mg of apratoxin A from the [methyl-^{13}C]methionine experiment. Approximately 2 mg of apratoxin A harvested from previous non-labeled batches was similarly obtained and used for NMR enrichment comparisons (Figure 5.2).

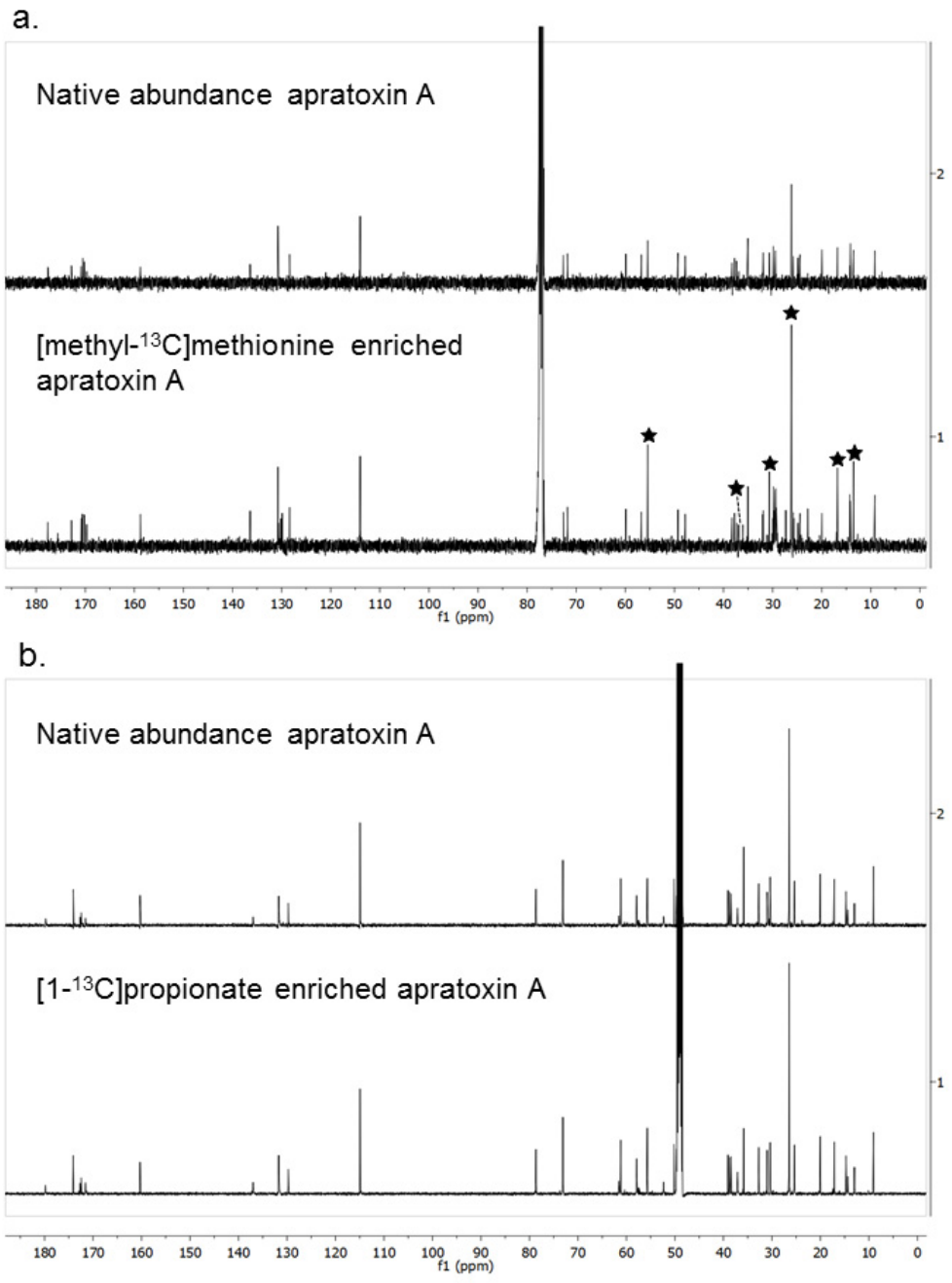


Figure 5.2: Comparison of ^{13}C-NMR enrichment of apratoxin A with methionine and propionate

a. Unlabeled apratoxin A on top compared with apratoxin A labeled with [methyl-^{13}C]methionine on bottom (in $CDCl_3$). Stars indicate enriched carbon resonances. **b.** Unlabeled apratoxin A on top compared with apratoxin A labeled with [1-^{13}C]propionate on bottom (in D_4-MeOH). There is no significant enrichment in any of the ^{13}C resonances.

NMR analysis and calculation of ^{13}C incorporation levels

All ^{13}C NMR analyses were performed on a Varian VX500 spectrometer at 500 MHz for ^{1}H NMR and 125 MHz for ^{13}C NMR by Nathan Moss (Tables 5.1- 5.3). A JEOL ECZ500 spectrometer was used to obtain HSQC and HMBC spectra (data not shown) in order to confirm $^{13}C/^{1}H$ shifts of apratoxin A in D_4-methanol solvent used in the analysis of the [1-^{13}C]propionate feeding experiment (Table 5.2). Residual solvent shifts in each sample were used as the chemical shift references. Sample processing was performed using MestReNova and Delta. Table 5.1 indicates ^{13}C-enrichment by carbon number in the [methyl-^{13}C]methionine feeding experiment, using a normalization method adapted from previous studies[23]. Briefly, integrated values of natural-abundance apratoxin A ^{13}C NMR signals are indicated in column 4. Five carbon atoms with different biosynthetic origins and chemical shifts were used for normalization between unenriched and enriched spectra: C7, C13, C20, C30, and C37. Normalization factors (columns 5-9) were obtained by dividing the integral values of all ^{13}C NMR signals in column 4 by those of C7, C13, C20, C30, and C37 (underlined in column 4). Then, the integral for the enriched signals of each of the five carbon atoms used in the normalization procedure (underlined in column 10) were multiplied by their normalization factors (columns 5-9) to obtain an expected signal integral for all carbon signals if no enrichment had occurred (columns 11-15). Percent enrichment was calculated by subtracting columns 11-15 from column 10, dividing the result by the column 11-15 values (*(enriched – expected unenriched)/expected unenriched*), and multiplying by 100 to yield the values in columns 16-20. The five results in columns 16-20 were averaged to yield the values in column 21. Column 22 is another expression of enrichment, namely the fold change of the enriched vs. unenriched sample.

Construct design

All constructs for apratoxin biosynthetic proteins were amplified from a plasmid encoding full length *aprA* (pAPS1)[119] and inserted into pMCSG7[166] by ligation independent cloning (LIC) to create expression plasmids for AprA ΨGNAT-MT2$_L$ (residues 502-1022, pAPS4), AprA ΔACP (residues 2-1034, pAPS5), and AprA MT2$_L$ (residues 628-1022, pMAS247). Primers are listed in Table 5.4. All *aprA* site-directed mutations were introduced into pMAS247 using the QuickChange protocol (Stratagene). Plasmid pMAS308 encoding BryX MT2$_L$

Table 5.1 Enrichment of apratoxin A by culture of *M. bouillonii* PNG 5-198 supplemented with [methyl-^{13}C]methionine

Column # → 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
			Integral of	Normalization factor using:					Integral of	Normalized integral of enriched sample using:					Enrichment compared to:					Average %	Fold
C#	Biosynthetic source	ppm	unlabeled resonance	C7	C13	C20	C30	C37	enriched resonance	C7	C13	C20	C30	C37	C7	C13	C20	C30	C37	Enrichment[1]	change
1	Pro	172.8	42.7	0.4	0.7	0.4	0.4	0.6	45.9	37.2	40.2	30.6	36.2	40.1	23.3	14.1	50.0	26.7	14.5	25.7	1.26
2	Pro	59.9	112.6	1.1	1.8	1.1	1.0	1.6	79.2	98.1	106.1	80.7	95.5	105.7	-19.3	-25.3	-1.9	-17.1	-25.1	-17.7	0.82
3	Pro	29.5	119.5	1.1	2.0	1.2	1.1	1.6	126.5	104.1	112.6	85.6	101.3	112.2	21.5	12.4	47.7	24.8	12.8	23.8	1.24
4	Pro	25.8	91.9	0.9	1.5	0.9	0.8	1.3	81.7	80.1	86.6	65.9	78.0	86.3	1.9	-5.7	24.0	4.8	-5.4	3.9	1.04
5	Pro	47.8	133.9	1.3	2.2	1.3	1.2	1.8	88.1	116.7	126.2	96.0	113.6	125.7	-24.5	-30.2	-8.2	-22.5	-30.0	-23.1	0.77
6	Ile	170.8	37.4	0.4	0.6	0.4	0.3	0.5	51.4	32.6	35.2	26.8	31.7	35.1	57.9	46.1	92.1	62.3	46.6	61.0	1.61
7	Ile	56.8	104.0	1.0	1.7	1.0	0.9	1.4	90.7	90.7	98.0	74.6	88.2	97.7	0.0	-7.5	21.6	2.8	-7.2	1.9	1.02
8	Ile	31.9	82.9	0.8	1.4	0.8	0.7	1.1	68.2	72.3	78.1	59.4	70.3	77.9	-5.7	-12.7	14.7	-3.1	-12.4	-3.8	0.96
9	Ile	24.8	93.7	0.9	1.5	0.9	0.8	1.3	66.0	81.7	88.3	67.2	79.5	88.0	-19.2	-25.3	-1.7	-17.0	-25.0	-17.6	0.82
10	Ile	9.2	124.2	1.2	2.0	1.2	1.1	1.7	83.4	108.3	117.0	89.0	105.4	116.6	-23.0	-28.8	-6.4	-20.9	-28.5	-21.5	0.79
11	Ile	14.2	145.1	1.4	2.4	1.4	1.3	2.0	86.1	126.5	136.7	104.0	123.1	136.3	-31.9	-37.0	-17.2	-30.1	-36.8	-30.6	0.69
12	SAM	30.7	89.0	0.9	1.5	0.9	0.8	1.2	175.7	77.6	83.8	63.8	75.5	83.6	126.5	109.6	175.5	132.8	110.3	***130.9***	***2.31***
13	Ala	170.2	61.1	0.6	1.0	0.6	0.5	0.8	57.6	53.3	57.6	43.8	51.8	57.4	8.1	0.0	31.5	11.1	0.3	10.2	1.10
14	Ala	60.9	76.9	0.7	1.3	0.8	0.7	1.1	90.6	67.0	72.5	55.1	65.2	72.2	35.1	25.0	64.3	38.9	25.5	37.8	1.38
15	Ala	14.1	62.4	0.6	1.0	0.6	0.6	0.9	70.6	54.4	58.8	44.7	52.9	58.6	29.9	20.1	57.9	33.5	20.6	32.4	1.32
16	SAM	36.9	38.8	0.4	0.6	0.4	0.3	0.5	109.7	33.8	36.6	27.8	32.9	36.4	224.3	200.0	294.4	233.2	201.0	***230.6***	***3.31***
17	Tyr	170.6	58.8	0.6	1.0	0.6	0.5	0.8	58.5	51.3	55.4	42.2	49.9	55.2	14.1	5.6	38.8	17.3	5.9	16.3	1.16
18	Tyr	50.6	97.8	0.9	1.6	1.0	0.9	1.3	13.8	85.2	92.1	70.1	82.9	91.8	-83.8	-85.0	-80.3	-83.4	-85.0	-83.5	0.17
19	Tyr	37.3	116.5	1.1	1.9	1.2	1.0	1.6	95.5	101.6	109.8	83.5	98.8	109.4	-6.0	-13.0	14.4	-3.4	-12.7	-4.1	0.96
20	Tyr	128.4	100.6	1.0	1.6	1.0	0.9	1.4	s72.1	87.7	94.8	72.1	85.3	94.5	-17.8	-23.9	0.0	-15.5	-23.7	-16.2	0.84
21/25	Tyr	130.8	293.1	2.8	4.8	2.9	2.6	4.0	237.2	255.4	276.1	210.0	248.6	275.2	-7.2	-14.1	12.9	-4.6	-13.8	-5.4	0.95
22/24	Tyr	114.0	254.6	2.4	4.2	2.5	2.3	3.5	225.7	221.9	239.9	182.5	215.9	239.0	1.7	-5.9	23.7	4.5	-5.6	3.7	1.04
23	Tyr	158.8	43.0	0.4	0.7	0.4	0.4	0.6	46.0	37.5	40.5	30.8	36.5	40.4	22.6	13.5	49.1	26.0	13.8	25.0	1.25
26	SAM	55.4	124.2	1.2	2.0	1.2	1.1	1.7	270.1	108.2	117.0	89.0	105.3	116.6	149.6	130.9	203.5	156.5	131.7	***154.5***	***2.55***
27	Acetate	169.7	44.5	0.4	0.7	0.4	0.4	0.6	50.1	38.8	41.9	31.9	37.7	41.7	29.2	19.6	57.2	32.8	20.0	31.8	1.32
28	Acetate	130.6	42.1	0.4	0.7	0.4	0.4	0.6	56.6	36.7	39.6	30.1	35.7	39.5	54.4	42.8	87.7	58.7	43.3	57.4	1.57
29	Cys	136.4	90.2	0.9	1.5	0.9	0.8	1.2	98.7	78.6	85.0	64.6	76.5	84.7	25.6	16.2	52.7	29.1	16.6	28.0	1.28
30	Cys	72.6	111.7	1.1	1.8	1.1	1.0	1.5	94.8	97.4	105.3	80.1	94.8	104.9	-2.7	-10.0	18.3	0.0	-9.7	-0.8	0.99
31	Cys	37.7	71.5	0.7	1.2	0.7	0.6	1.0	37.6	62.3	67.4	51.3	60.7	67.2	-39.7	-44.2	-26.7	-38.0	-44.0	-38.5	0.62
32	SAM	13.5	81.2	0.8	1.3	0.8	0.7	1.1	167.6	70.8	76.5	58.2	68.9	76.2	136.8	119.1	188.0	143.4	119.8	***141.4***	***2.41***
33	Acetate	177.6	57.3	0.6	0.9	0.6	0.5	0.8	65.3	50.0	54.0	41.1	48.6	53.8	30.8	21.0	59.0	34.4	21.4	33.3	1.33
34	Acetate	49.3	76.3	0.7	1.2	0.8	0.7	1.1	84.9	66.5	71.9	54.7	64.7	71.7	27.6	18.0	55.2	31.1	18.5	30.1	1.30
35	Acetate	71.8	116.2	1.1	1.9	1.2	1.0	1.6	122.4	101.3	109.5	83.3	98.6	109.1	20.8	11.8	46.9	24.1	12.1	23.2	1.23
36	Acetate	38.3	94.7	0.9	1.6	0.9	0.8	1.3	104.0	82.6	89.3	67.9	80.3	89.0	25.9	16.5	53.1	29.4	16.9	28.4	1.28
37	Acetate	24.4	72.6	0.7	1.2	0.7	0.6	1.0	68.4	63.3	68.4	52.0	61.6	68.2	8.0	-0.1	31.4	11.0	0.3	10.1	1.10
38	Acetate	37.8	111.1	1.1	1.8	1.1	1.0	1.5	129.8	96.8	104.7	79.6	94.2	104.3	34.0	24.0	63.0	37.7	24.4	36.6	1.37
39	Acetate	77.4	N.P.*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
40	Acetate	35.0	93.4	0.9	1.5	0.9	0.8	1.3	94.2	81.4	88.0	66.9	79.2	87.7	15.8	7.1	40.8	19.0	7.5	18.0	1.18
41-43	SAM	26.2	387.6	3.7	6.3	3.9	3.5	5.3	659.6	337.8	365.2	277.8	328.7	363.9	95.2	80.6	137.4	100.6	81.2	***99.0***	***1.99***
44	SAM	16.8	94.1	0.9	1.5	0.9	0.8	1.3	203.0	82.0	88.6	67.4	79.8	88.3	147.5	129.0	201.0	154.4	129.8	***152.3***	***2.52***
45	Acetate	20.0	109.2	1.0	1.8	1.1	1.0	1.5	70.7	95.2	102.9	78.3	92.6	102.5	-25.7	-31.2	-9.6	-23.6	-31.0	-24.2	0.76

*N.P. = No quantitation due to $CDCl_3$ peak overlap

[1]Final enrichment values are derived by normalization of enriched ^{13}C-NMR resonances to five carbon atoms. Calculation steps are detailed in Methods section. SAM-derived enrichment values are in bold and italicized in column 21 and 22.

Table 5.2 Enrichment of ^{13}C in apratoxin A by culture of *M. bouillonii* PNG 5-198 in media supplemented with [1-^{13}C]propionate

Column # →																						
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
	Biosynthetic		Integral of unlabeled	Normalization factor using:					Integral of enriched	Normalized integral of enriched sample using:					Enrichment compared to:					Average % Enrichment[3]	Fold change
C#	source	ppm	resonance	C7	C13	C20	C30	C37	resonance	C7	C13	C20	C30	C37	C7	C13	C20	C30	C37		
1	Pro	174.1	45.5	0.5	0.9	0.9	0.2	0.3	56.0	44.5	63.0	50.2	50.6	52.1	25.8	-11.2	11.6	10.7	7.4	8.8	1.1
2	Pro	61.2	91.4	0.9	1.8	1.7	0.4	0.5	97.9	89.6	126.8	100.9	101.8	104.9	9.3	-22.8	-3.0	-3.8	-6.6	-5.4	0.9
3	Pro	29.8	2.2	0.0	0.0	0.0	0.0	0.0	2.9	2.1	3.0	2.4	2.4	2.5	37.2	-3.1	21.8	20.8	17.2	18.8	1.2
4	Pro	25.9	1.6	0.0	0.0	0.0	0.0	0.0	7.5	1.5	2.2	1.7	1.7	1.8	389.0	245.4	334.0	330.4	317.7		1.0
5[1]	Pro	N.P.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	Ile	172.7	32.0	0.3	0.6	0.6	0.2	0.2	38.7	31.4	44.4	35.3	35.6	36.7	23.5	-12.8	9.6	8.7	5.5	6.9	1.1
7	Ile	57.9	97.8	1.0	1.9	1.8	0.5	0.6	95.8	95.8	135.6	107.9	108.8	112.1	0.0	-29.4	-11.3	-12.0	-14.6	-13.4	0.9
8	Ile	31.0	65.8	0.7	1.3	1.2	0.3	0.4	79.9	64.4	91.2	72.6	73.2	75.4	24.0	-12.4	10.1	9.2	5.9	7.4	1.1
9	Ile	25.3*	173.2	1.8	3.4	3.2	0.8	1.0	198.6	169.6	240.2	191.1	192.7	198.6	17.1	-17.3	3.9	3.0	0.0	1.3	1.0
10	Ile	9.1	79.2	0.8	1.6	1.5	0.4	0.5	91.9	77.6	109.9	87.4	88.2	90.8	18.5	-16.3	5.1	4.3	1.2	2.5	1.0
11	Ile	14.7	73.3	0.7	1.4	1.4	0.4	0.4	85.5	71.8	101.6	80.9	81.6	84.0	19.1	-15.9	5.7	4.8	1.7	3.1	1.0
12	SAM	30.4	93.6	1.0	1.8	1.8	0.5	0.5	91.5	91.7	129.8	103.3	104.2	107.4	-0.2	-29.6	-11.5	-12.2	-14.8	-13.7	0.9
13	Ala	171.6	51.0	0.5	1.0	1.0	0.2	0.3	70.7	50.0	70.7	56.3	56.8	58.5	41.6	0.0	25.7	24.6	20.9	22.6	1.2
14	Ala	61.6	58.7	0.6	1.2	1.1	0.3	0.3	71.3	57.5	81.4	64.8	65.3	67.3	24.0	-12.4	10.1	9.1	5.9	7.4	1.1
15	Ala	14.4	64.4	0.7	1.3	1.2	0.3	0.4	70.6	63.1	89.3	71.1	71.7	73.9	11.8	-21.0	-0.7	-1.6	-4.5	-3.2	1.0
16	SAM	35.8	82.8	0.8	1.6	1.6	0.4	0.5	86.3	81.1	114.8	91.4	92.2	95.0	6.5	-24.8	-5.5	-6.3	-9.1	-7.9	0.9
17	Tyr	172.4	43.3	0.4	0.8	0.8	0.2	0.3	51.2	42.4	60.1	47.8	48.2	49.7	20.6	-14.8	7.0	6.2	3.0	4.4	1.0
18	Tyr	52.3	61.2	0.6	1.2	1.1	0.3	0.4	84.0	60.0	84.9	67.6	68.1	70.2	40.1	-1.0	24.3	23.3	19.7	21.3	1.2
19	Tyr	37.1	130.5	1.3	2.6	2.4	0.6	0.8	155.9	127.8	181.0	144.0	145.2	149.6	22.0	-13.9	8.2	7.3	4.2	5.6	1.1
20	Tyr	129.7	53.3	0.5	1.0	1.0	0.3	0.3	58.9	52.2	74.0	58.9	59.4	61.2	12.7	-20.4	0.0	-0.8	-3.8	-2.5	1.0
21/25	Tyr	131.7	214.8	2.2	4.2	4.0	1.0	1.2	243.0	210.4	297.9	237.1	239.1	246.3	15.5	-18.4	2.5	1.6	-1.4	0.0	1.0
22/24	Tyr	115.0	206.4	2.1	4.0	3.9	1.0	1.2	215.3	202.1	286.2	227.8	229.7	236.7	6.5	-24.8	-5.5	-6.3	-9.0	-7.8	0.9
23	Tyr	160.3	50.0	0.5	1.0	0.9	0.2	0.3	48.6	48.9	69.3	55.2	55.6	57.3	-0.7	-29.9	-11.9	-12.7	-15.2	-14.1	0.9
26	SAM	55.7	100.0	1.0	2.0	1.9	0.5	0.6	100.0	97.9	138.7	110.4	111.3	114.7	2.1	-27.9	-9.4	-10.2	-12.8	-11.6	0.9
27	Acetate	171.3	4.8	0.0	0.1	0.1	0.0	0.0	4.6	4.7	6.6	5.3	5.3	5.5	-1.5	-30.5	-12.6	-13.4	-15.9	-14.8	0.9
28[2]	Acetate	N.P.	8.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29	Cys	137.0	58.0	0.6	1.1	1.1	0.3	0.3	80.2	56.8	80.5	64.0	64.6	66.5	41.2	-0.3	25.3	24.2	20.6	22.2	1.2
30	Cys	73.1*	207.4	2.1	4.1	3.9	1.0	1.2	230.8	203.1	287.6	228.9	230.8	237.8	13.6	-19.7	0.8	0.0	-2.9	-1.6	1.0
31	Cys	38.4	76.2	0.8	1.5	1.4	0.4	0.4	94.2	74.7	105.7	84.1	84.9	87.4	26.2	-10.9	12.0	11.0	7.8	9.2	1.1
32	SAM	13.0	72.5	0.7	1.4	1.4	0.3	0.4	87.3	71.0	100.5	80.0	80.7	83.1	23.0	-13.1	9.2	8.3	5.1	6.5	1.1
33	Acetate	179.8	43.7	0.4	0.9	0.8	0.2	0.3	43.7	42.8	60.6	48.2	48.6	50.1	2.1	-27.9	-9.4	-10.2	-12.8	-11.6	0.9
34	Acetate	50.2	76.2	0.8	1.5	1.4	0.4	0.4	92.0	74.6	105.6	84.1	84.8	87.4	23.3	-12.9	9.4	8.5	5.3	6.7	1.1
35	Acetate	73.1*	207.4	2.1	4.1	3.9	1.0	1.2	230.8	203.1	287.6	228.9	230.8	237.8	13.6	-19.7	0.8	0.0	-2.9	-1.6	1.0
36	Acetate	39.1	79.8	0.8	1.6	1.5	0.4	0.5	94.7	78.2	110.7	88.1	88.8	91.5	21.1	-14.5	7.5	6.6	3.4	4.8	1.0
37	Acetate	25.3*	173.2	1.8	3.4	3.2	0.8	1.0	198.6	169.6	240.2	191.1	192.7	198.6	17.1	-17.3	3.9	3.0	0.0	1.3	1.0
38	Acetate	38.7	77.6	0.8	1.5	1.5	0.4	0.4	82.8	76.0	107.6	85.6	86.4	89.0	9.0	-23.0	-3.3	-4.1	-6.9	-5.7	0.9
39	Acetate	78.7	87.4	0.9	1.7	1.6	0.4	0.5	102.0	85.6	121.2	96.5	97.3	100.2	19.2	-15.9	5.7	4.9	1.8	***3.1***	***1.0***
40	Acetate	32.7	93.5	1.0	1.8	1.8	0.5	0.5	93.8	91.6	129.7	103.2	104.1	107.3	2.3	-27.7	-9.2	-9.9	-12.6	-11.4	0.9
41-43	SAM	26.5	404.2	4.1	7.9	7.6	1.9	2.3	487.6	395.8	560.5	446.1	449.8	463.5	23.2	-13.0	9.3	8.4	5.2	6.6	1.1
44	SAM	17.1	88.5	0.9	1.7	1.7	0.4	0.5	95.7	86.7	122.7	97.7	98.5	101.5	10.4	-22.0	-2.0	-2.8	-5.7	-4.4	1.0
45	Acetate	20.0	89.7	0.9	1.8	1.7	0.4	0.5	103.2	87.8	124.4	99.0	99.8	102.8	17.5	-17.0	4.2	3.4	0.3	1.7	1.0

*Overlapping ^{13}C-NMR shifts. The enrichment values cannot be distinguished, so both are reported.

[1]C5 – N.P. = No quantitation. C5 obscured by D_4-MeOH solvent peak.

[2]C28 – N.P. = Weak signal observed in native abundance, but not enriched sample.

[3]Final enrichment values are derived by normalization of enriched ^{13}C-NMR resonances to five different carbon atoms. Calculation steps are detailed in methods. C39 enrichment value is italicized and bolded in columns 21 and 22.

Table 5.3 ^{13}C-NMR shifts of native-abundance apratoxin A

	[methyl-^{13}C] methionine labeled CDCl3	Native abundance $CDCl_3$	[1-^{13}C]propionate-labeled D_4-MeOH	Native abundance D_4-MeOH
C#	ppm	ppm	ppm	ppm
1	172.8	172.8	174.1	174.1
2	59.9	59.9	61.2	61.2
3	29.5	29.4	29.8	29.8
4	25.8	25.8	25.9	25.9
5	47.8	47.8	47.5	47.5
6	170.8	170.8	172.7	172.7
7	56.8	56.7	57.9	57.9
8	31.9	31.9	31.0	31.0
9	24.8	24.8	25.3	25.3
10	9.2	9.2	9.1	9.1
11	14.2	14.2	14.7	14.7
12	30.7	30.7	30.4	30.4
13	170.2	170.2	171.6	171.6
14	60.9	60.8	61.6	61.6
15	14.1	14.1	14.4	14.4
16	36.9	36.9	35.8	35.8
17	170.6	170.5	172.4	172.4
18	50.6	50.6	52.3	52.3
19	37.3	37.4	37.1	37.1
20	128.4	128.4	129.7	129.7
21/25	130.8	130.8	131.7	131.7
22/24	114.0	114.0	115.0	115.0
23	158.8	158.8	160.3	160.3
26	55.4	55.4	55.7	55.7
27	169.7	169.7	171.3	171.3
28	130.6	130.6	N.P.	130.8
29	136.4	136.4	137.0	137.0
30	72.6	72.6	73.1	73.1
31	37.7	37.7	38.4	38.4
32	13.5	13.5	13.0	13.0
33	177.6	177.6	179.8	179.8
34	49.3	49.3	50.2	50.2
35	71.8	71.8	73.1	73.1
36	38.3	38.3	39.1	39.1
37	24.4	24.4	25.3	25.3
38	37.8	37.8	38.7	38.7
39	N.P.*	N.P.	78.7	78.7
40	35.0	35.0	32.7	32.7
41/42/43	26.2	26.2	26.5	26.5
44	16.8	16.8	17.1	17.1
45	20.0	20.0	20.0	20.0

^{13}C-NMR shifts of native-abundance apratoxin A in $CDCl_3$ and D_4-MeOH, [methyl-^{13}C]methionine labeled apratoxin A in $CDCl_3$, and [1-^{13}C]propionate-labeled apratoxin A in D_4-MeOH.

Table 5.4 Primers for cloning $MT2_L$ and FabD constructs

AprA ΨGNAT-$MT2_L$ (502-1022)	pAPS4_F	**TACTTCCCATCCAATGCA**GAAAAGCGCAAATATCAGATACGGTATGCAACT
	pAPS4_R	**TTATCCACTTCCAATGTTA**TGTTTTCTCCTGTACTTGATGC
AprA ΔACP (2-1034)	pAPS5_F	**TACTTCCAATCCAATGCA**CTAGATAAAATAAATCGTTATGCTCATGGGTTTGTAGC
	pAPS5_R	**TTATCCACTTCCAATGTTA**ATTTTCTGTCAAATTTTGTTTAATTTTTTC
AprA $MT2_L$ (628-1022)	pMAS247_F	**TACTTCCAATCCAATGCGCA**ATCTTCAGTTTTAAACAAAAAGCTTATAG
	pMAS247_R	**TTATCCACTTCCAATGTTA**TGTTTTCTCCTGTACTTGATGC
FabD	pMAS337_F	**TACTTCCAATCCAATGCC**ATGATAAAGACTGCATGGGTGTTTCCCG
	pMAS337_R	**TTATCCACTTCCAATGCTA**CCTTACTACTAAGCATTGCTGCAGATCAG
AprI ACP (2085-2201)	pMAS373_F	**TACTTCCAATCCAATGCC**TCAATAGCTACAAAGGATAATCTCCTTTTAGAAC
	pMAS373_R	**TTATCCACTTCCAATGCTA**TAGCTCTACCTCTATCCAGTTATTTTCTTTAATG
BryX $MT2_L$ (656-1052)	pMAS308_F	**TACTTCCAATCCAATGCC**ATGGAAAGTGAGTCAGTTGATGTT
	pMAS308_R	**TTATCCACTTCCAATGCTA**ATTACTAGACAGGTCTTCATTAATCC
AprA $MT2_L$ F797Y	pMAS310_ F	GATATGGAGCTATTTGCTGGACTATATCTAGGACATCGT
	pMAS310_R	ACGATGTCCTAGATATAGTCCAGCAAATAGCTCCATATC
AprA $MT2_L$ H914N	pMAS260	ATGTGGTAATTGCTAATAACGTACTCAATAACACAAAATTAATTCATCAAACC
AprA $MT2_L$ N915A	pMAS252	GATGTGGTAATTGCTAATAACGTACTCCATGCCACAAAATTAATTCATCAAACCTTAAATAA
AprA $MT2_L$ E940A	pMAS312_ F	GGGGGTTATTGGCATTACTAGCGTTTACTCAACCAATTGATAT
	pMAS312_R	ATATCAATTGGTTGAGTAAACGCTAGTAATGCCAATAACCCCC
AprA $MT2_L$ G952F	pMAS311_ F	CTAGAGTTTACTCAACCAATTGATATTCTTTTATACTTTTTCGGGTTGCTTCAAGGATTTTG
	pMAS311_R	CAAAATCCTTGAAGCAACCCGAAAAAGTATAAAAGAATATCAATTGGTTGAGTAAACTCTAG
AprA $MT2_L$ F958W	pMAS381_ F	TACTTTGGAGGGTTGCTTCAAGGATGGTGGTTGTTTGAAGATCC
	pMAS381_R	GGATCTTCAAACAACCACCATCCTTGAAGCAACCCTCCAAAGTA
AprA $MT2_L$ C972P	pMAS315_ F	GAATACCGACTAGAAGTTGGTCCTTTACTGAGTATACCACTGTG
	pMAS315_R	CACAGTGGTATACTCAGTAAAGGACCAACTTCTAGTCGGTATTC
AprA $MT2_L$ C972S	pMAS319_ F	AATACCGACTAGAAGTTGGTAGTTTACTGAGTATACCACTG
	pMAS319_R	CAGTGGTATACTCAGTAAACTACCAACTTCTAGTCGGTATT

All sequences are 5'- 3'. Bold font indicates handles for ligation-independent insertion into expression vector.

(residues 656-1052) was amplified from a *Candidatus Endobugula sertula* cosmid library[20] and inserted into pMCSG7 via LIC. DNAs encoding FabD and AprI ACP (residues 2085-2201) were amplified from *Moorea bouillonii* PNG 5-198 genomic DNA[184] and inserted into pMCSG7 via LIC to produce pMAS337 and pMAS373, respectively. All constructs and mutations were verified by Sanger sequencing at the University of Michigan DNA Sequencing Core.

Protein expression and purification

AprA apo-ACP and AprA MT_L-ΨGNAT were expressed and purified as described previously[119]. In order to produce AprA holo-ACP and AprI holo-ACP, plasmids were transformed into *Escherichia coli* strain BAP1[220] and expressed and purified identically to AprA ACP. pAPS4 (AprA ΨGNAT- $MT2_L$), pMAS247 (AprA $MT2_L$), and pMAS337 (FabD) were transformed into *Escherichia coli* strain BL21(DE3), grown in 0.5 L of TB media with 100 μg mL^{-1} ampicillin to an OD_{600} = 1-2 at 37°C. Cultures were cooled to 20°C, induced with 200 μM IPTG, and expressed overnight. To produce AprA ΔACP, pAPS5 was transformed into BL21(DE3) containing the pRare-CDF plasmid[48] and the pG-KJE8 plasmid (Takara) encoding the DnaK, DnaJ, GrpE, GroEL and GroES chaperones. Cells were grown in 0.5 L of TB media with 100 μg mL^{-1} ampicillin, 50 μg mL^{-1} spectinomycin, and 35 μg mL^{-1} chloramphenicol at 37°C. *dnaK*, *dnaJ*, and *grpE* were induced with 0.2 % (w/v) arabinose at OD_{600} = ~1. Cultures were grown for an additional 30 minutes at 37°C, cooled to 20°C for 1 hour, induced with 120 μM IPTG, and expressed overnight. AprA $MT2_L$ and BryX $MT2_L$ cell pellets were resuspended in 35 mL Tris buffer A (50 mM Tris pH 7.4, 300 mM NaCl, 10% (v/v) glycerol, 20 mM imidazole) with 0.1 mg mL^{-1} lysozyme, 0.05 mg mL^{-1} DNase, and 2 mM $MgCl_2$, incubated on ice for 30 min, lysed by sonication, and cleared by centrifugation (38,650 x g, 30 min). Supernatant was filtered with a 0.45 μm syringe filter, loaded onto a 5 mL HisTrap column (GE Healthcare), washed with 10 column volumes of Tris buffer A, and eluted with a 5-100% linear gradient of Tris buffer B (50 mM Tris pH 7.4, 300 mM NaCl, 10% (v/v) glycerol, 400 mM imidazole). Proteins were further purified by size exclusion chromatography (HiLoad 16/60 Superdex S200) in Tris buffer C (50 mM Tris pH 7.4, 150 mM NaCl, 10% (v/v) glycerol).

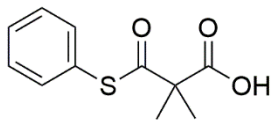
FabD was purified identically to AprA $MT2_L$ and BryX $MT2_L$ with the addition of 100 mM urea to the lysis buffer. Following size-exclusion chromatography, FabD was diluted into Tris buffer D (50 mM Tris pH 7.4, 20 mM NaCl, 10% glycerol) and further purified by anion exchange.

The protein was loaded onto a 5 mL HiTrap Q column (GE Healthcare), washed with 10 column volumes of Tris Buffer E (50 mM Tris pH 7.4, 10% glycerol), and eluted with a 0-100 % gradient of Tris buffer F (50 mM Tris pH 7.4, 1M NaCl, 10% glycerol) over 40 column volumes, and dialyzed overnight into Tris buffer C.

AprA ΨGNAT- $MT2_L$ cell pellets were resuspended in 5 mL Tris buffer G (100 mM Tris pH 7.9, 500 mM NaCl, 5% glycerol, 15 mM imidazole) / g of cell paste. Resuspended cell pellets were incubated on ice with 4 mg DNase, 10 mg lysozyme, and 4 mM $MgCl_2$ for 30 min. Cells were lysed by three passes through an Avestin EmulsiFlex-C3 homogenizer and clarified by centrifugation at 30,000 x g. The supernatant was loaded onto a 5mL HisTrap column and eluted with a 10-100% gradient of Tris buffer H (100 mM Tris pH 7.9, 500 mM NaCl, 5% glycerol, 300 mM imidazole). The His-tag was removed by overnight incubation with tobacco etch virus protease in Tris buffer I (50 mM Tris pH 7.9, 100 mM NaCl, 10% glycerol) followed by a second 5-mL HisTrap column. AprA ΨGNAT-$MT2_L$ was further purified via size exclusion chromatography (HiLoad 16/600 Superdex 200) in Tris buffer J (50 mM Tris pH 7.9, 100 mM NaCl).

AprA ΔACP cell pellets were resuspended in 5 mL HEPES buffer A (50 mM HEPES pH 7, 50 mM $(NH_4)_2SO_4$, 10 % glycerol, 15 mM imidazole) / g of cell paste. Resuspended cell pellets were incubated on ice with 4 mg DNase, 10 mg lysozyme, and 4 mM $MgCl_2$ for 30 min. Cells were lysed by three passes through an Avestin EmulsiFlex-C3 homogenizer and clarified by centrifugation at 30,000 x g. The supernatant was loaded onto a 5 mL HisTrap column and eluted with a 10-100% gradient of HEPES buffer B (50 mM HEPES pH 7, 50 mM $(NH_4)_2SO_4$, 10 % glycerol, 300 mM imidazole). AprA ΔACP was further purified via size exclusion chromatography (HiLoad 16/600 Superdex 200) in HEPES Buffer C (50 mM HEPES pH 7, 50 mM $(NH_4)_2SO_4$, 10 % glycerol). The sample was buffer exchanged into 50 mM HEPES pH 7, 100 mM NaCl via size exclusion chromatography (HiLoad 16/600 Superdex 200) prior to EM analysis.

Dimethylmalonyl-thiophenol synthesis



All synthesis was performed by Andrew Lowell. 2,2-Dimethyl-3-oxo-3-(phenylthiol)propanoic acid. The following reaction was adapted from a published procedure[101]. Dimethyl malonic acid (1.36 g, 10.3 mmol) and thiophenol (1.06 mL, 10.3 mmol) were dissolved in acetonitrile (120 mL) and stirred in an ice bath. Diisopropylcarbodiimide (1.01 mL, 6.52 mmol) was slowly added and the mixture was warmed to room temperature overnight with stirring. The reaction was quenched by the addition of saturated sodium bicarbonate (500 mL), filtered, and extracted with Et_2O (2 x 250 mL). The aqueous layer was carefully acidified by the addition of 4 M HCl (pH = 2) and extracted with Et_2O (3 x 250 mL). The organic layers used to extract the acidic aqueous layer were combined, washed with brine (500 mL), dried with Na_2SO_4, and concentrated to a colorless oil. The oil was recrystallized from Et_2O/hexanes to give a mixture (0.329 g) of 2,2-dimethyl-3-oxo-3-(phenylthio)propanoic acid and diisopropyl urea as a white crystalline solid. The mother liquor was concentrated to a colorless oil and purified using a flash chromatography system (16-100% EtOAc/hexanes, SiliaSep HP 12 g, 36 mL min^{-1}) to give an amorphous white solid. This residue was purified using a flash chromatography system (16-25% EtOAc/hexanes, SiliaSep HP 12 g, 36 mL min^{-1}) to give 2,2-dimethyl-3-oxo-3-(phenylthio)propanoic acid (0.204 g, 13.9%) as an amorphous white powder: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.42 (s, 5H), 1.62 (s, 6H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 198.5, 177.4, 135.0, 129.9, 129.5, 126.8, 57.3, 23.3; HRMS (ES) calcd for $C_{11}H_{12}O_3SNa$ (MNa^+) 247.0399, found 247.0397.

Production of acyl-ACPs

In order to produce malonyl- (Mal-), methylmalonyl- (MeMal-), and acetoacetyl- (AcAc-) ACP, 180 μM AprA apo-ACP was incubated with 20 μM *Streptomyces verticillus* phosphopantetheinyl transferase (SVP)[142], 20 mM $MgCl_2$, and 3-4 fold molar excess of the corresponding CoA for 4 hr at 30°C in Tris buffer C. Acyl-ACPs were purified from reaction mixtures using size exclusion chromatography (HiLoad 16/60 Superdex S75) equilibrated with 100 mM Tris pH 7.4, 250 mM NaCl, 5% glycerol, 5 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP).

Dimethylmalonyl- (Me_2Mal-)ACP for intact protein spectra was prepared by incubating 300 μM AprA MeMal-ACP with 150 μM AprA MT_L, 3 mM $(NH_4)_2Fe(SO_4)_2$, and 6 mM SAM in 50 mM HEPES pH 7.4, 150 mM NaCl. Reaction mixtures (500 μL) were incubated at 30°C for 5.5 hours. Me_2Mal-ACP was isolated from the reaction mixture via size exclusion chromatography

(HiLoad 16/600 Superdex 75) with 100 mM Tris pH 7.4, 250 mM NaCl, 5% glycerol, 5 mM TCEP. Me_2Mal-ACP for Ppant ejection experiments was prepared by incubating 100 μM AprA holo-ACP with 25 mM Me_2Mal-thiophenol in 300 mM sodium bicarbonate pH 8.1 at 25°C overnight. Excess thiophenol was removed via size exclusion chromatography (HiLoad 16/600 Superdex 75) with 100 mM Tris pH 7.4, 250 mM NaCl, 5% glycerol, 5 mM TCEP.

AprA $MT2_L$ enzyme assays

Reaction mixtures (10 μL) containing 100 μM Me_2Mal-ACP, 25 μM AprA $MT2_L$, and 0.5 mM SAM, SAH, or sinefungin in 50 mM HEPES pH 7.4, 150 mM NaCl, 0.5 mM $MgCl_2$ were incubated at 30°C for 1 min. Reactions were quenched with 10% (v/v) formic acid, and 0.25 μL of reaction mixtures were subjected to LC-MS analysis.

Reaction mixtures (10 μL) containing 100 μM MeMal-ACP, 25 μM AprA $MT2_L$, and 0.5 mM SAM, SAH, or sinefungin in 50 mM HEPES pH 7.4, 150 mM NaCl, 0.5 mM $MgCl_2$ were incubated at 30°C for 10 min. Reactions were quenched with 10% (v/v) formic acid, and 0.25 μL of reaction mixtures were subjected to LC-MS analysis.

Reaction mixtures (10 μL) containing 100 μM Mal-ACP, 25 μM AprA $MT2_L$, and 0.5 mM SAM, SAH, or sinefungin in 50 mM HEPES pH 7.4, 150 mM NaCl, 0.5 mM $MgCl_2$ were incubated at 30°C for 2 hr. Reactions were quenched with 10% (v/v) formic acid, and 0.2 μL of reaction mixtures were subjected to LC-MS analysis.

Reaction mixtures for intact protein analysis (10 μL) containing 100 μM AcAc-ACP, 25 μM AprA $MT2_L$, and 0.5 mM SAM in 50 mM HEPES pH 7.4, 150 mM NaCl, 0.5 mM $MgCl_2$ were incubated at 30°C for 15 min. Reaction mixtures for Ppant ejection analysis of amino acid variants (15 μL) containing 50 μM AcAc-ACP, 12.5 μM AprA $MT2_L$, and 0.5 mM SAM in 50 mM HEPES pH 7.4, 150 mM NaCl, 0.5 mM $MgCl_2$ were incubated at 30°C for 15 min. Reactions were quenched with 10% (v/v) formic acid, and 0.1 μL of reaction mixtures were subjected to LC-MS analysis.

Reactions (10 μL) to test AprA $MT2_L$ for acyltransfer activity contained 100 μM holo-ACP, 15 μM AprA $MT2_L$, 1 mM SAM, and 1 mM Mal-CoA in 50 mM HEPES pH 7.4, 150 mM NaCl were incubated at 30° C for 1 hr. Reactions were quenched with 10% (v/v) formic acid, and 0.1 μL of reaction mixtures were subjected to LC-MS analysis.

FabD enzyme assays

Reaction mixtures (120 μL) containing 100 μM AprA holo-ACP or AprI holo-ACP were incubated with 25 nM FabD and 0.85 mM Mal-CoA in 50 mM HEPES 7.4, 150 mM NaCl. Reactions were quenched (10 μL) with 10% (v/v) formic acid during a 10 min time course, and 1 μL of reaction mixtures were subjected to LC-MS analysis.

LC-MS analysis

Assay reaction mixtures were analyzed using a simultaneous intact protein and phosphopantetheine (Ppant) ejection method[150, 169] on an Agilent Q-TOF 6545. Samples were subjected to reverse phase HPLC (Phenomenex Aeris widepore C4 column 3.6 μM, 50 x 2.10 mm) at a flow rate of 0.5 mL min^{-1} in H_2O with 0.2% (v/v) formic acid. Protein was eluted over 4 minutes with a gradient of 5-100% acetonitrile with 0.2% (v/v) formic acid. Reactions containing $MT2_L$ and AcAc-ACP and the FabD enzyme assays were subjected to the following conditions: fragmentor voltage, 300 V; skimmer voltage, 75 V; nozzle voltage, 1000 V; sheath gas temperature, 350 °C; drying gas temperature, 325 °C. In order to limit in-source decay of Mal-ACP, MeMal-ACP, and Me_2Mal-ACP, samples from corresponding reactions with AprA $MT2_L$ were analyzed using: fragmentor voltage, 225 V; skimmer voltage, 25 V; nozzle voltage, 1000 V; sheath gas temperature, 350 °C; drying gas temperature, 325 °C.

Data were processed using MassHunter Qualitative Analysis Software (Agilent). The maximum entropy deconvolution algorithm was used to obtain intact protein masses. The relative abundances of Ppant ejection fragments were used to calculate fractions of the ACP species. For FabD malonyl loading assays, the abundances of Mal-ACP and its in-source decay product acetyl-ACP were combined to calculate the amount of Mal-ACP produced.

HPLC analysis

In order to analyze SAM consumption, 25 μL reaction mixtures containing 500 μM Mal-ACP, MeMal-ACP, or AcAc-ACP were incubated with 125 μM AprA $MT2_L$ and 0.5 mM SAM in 50 mM HEPES pH 7.4, 150 mM NaCl, 0.5 mM $MgCl_2$. Reactions were incubated for 10 min (MeMal-ACP), 15 min (AcAc-ACP), or 2 hr (Mal-ACP) at 30°C and quenched with 50 μL methanol. Precipitated protein was removed by centrifugation.

Reverse phase HPLC (Phenomenex Luna C18 column 5 μM, 250 x 4.6 mm) was used to separate small molecule components of the reaction (10 μL reaction mixtures, 2 μL of 1 mM SAM or 1 mM SAH standards) using HPLC buffer A (100 mM NaH_2PO_4, 75 mM NaOAc, pH 4.6) and HPLC buffer B (70% buffer A, 30% methanol) at a flow rate of 1 mL min^{-1}. Samples were injected onto the column equilibrated in 10% buffer B. Reaction components were eluted with a linear gradient to 30% buffer B (9% methanol) over 23 minutes. Absorbance was monitored at 254 nm.

Protein crystallization and structure determination

All crystallization and structure determination was performed by Andrew Sikkema and Rebecca Sturgis. AprA ΨGNAT-MT2$_L$ (503-1022) was crystallized by vapor diffusion in a 0.75:0.75 μL mixture of protein stock (5 mg mL^{-1} AprA in Tris buffer I with 1 mM SAM) and reservoir solution (0.01-0.05M trimethylamine *N*-oxide, 12-17% PEG 8000, 0.12 M Tris pH 7.5) in sitting drops at 20°C. Diffraction quality crystals obtained through micro seeding grew in 2-4 days. Crystals were cryoprotected with reservoir solution supplemented with 10% glycerol, and flash cooled in liquid N_2.

Diffraction data were collected at 100 K on GM/CA beamline 23ID-B at the Advanced Photon Source (APS) at Argonne National Laboratory (Argonne, IL). Data were processed using XDS (Table 5.5)[128]. The structure of AprA ΨGNAT-MT2$_L$ with bound SAH was solved by molecular replacement with Phaser[172] in the PHENIX software suite[130] using search models created with Sculptor[221] from the CurJ CMT (33% identity, PDB code: 5THY)[117] and CurA GNAT (24% identity, PDB code: 2REE)[31]. The initial model was modified with AutoBuild[171] to generate a 77% complete model of the AprA ΨGNAT-MT2$_L$. The remaining portion of the model was completed manually using Coot[132]. Refinement was performed using phenix.refine (Table 5.5)[173]. The structure was validated with MolProbity[133]. Figures were prepared using PyMOL[136]. Sequence alignments were prepared using Clustal[137] through Jalview[138]. The dendrogram was created using the average distance BLOSUM62 algorithm in Jalview.

Table 5.5 AprA $MT2_L$ crystallographic information

Data Collection	
Space group	*C*2
Cell dimensions a,b,c (Å)	152.4, 54.2, 109.3
α,β,γ (°)	90, 131.8, 90
X-ray source	APS 23ID-B
Wavelength (Å)	1.0332
d_{min} (Å)	2.25 (2.38-2.25)[1]
R_{merge}	0.090 (0.998)
Wilson B factor	53.0
Avg I/σ(I)	10.87 (1.53)
Completeness (%)	98.4 (97.8)
Multiplicity	6.1 (6.4)
Total observations	191,196 (31,888)
$CC_{1/2}$	0.998 (0.720)
CC*	0.999 (0.915)
Refinement	
Data range (Å)	48.86-2.25
Reflections in refinement (#)	31,507
R_{work}/R_{free}	17.4/23.0
Non-hydrogen atoms (#)	4,317
protein	4,161
ligand	31
water	125
Amino acid residues	511
Deviation from ideality	
bond lengths (Å)	0.008
bond angles (°)	0.983
Average B-factor	68.1
macromolecule	68.3
ligand	96.2
solvent	62.2
Ramachandran plot	
favored (%)	95.9
allowed (%)	3.9
outliers (%)	0.2
PDB	6D6Y

[1] Values in parentheses pertain to outermost shell of data.

Negative-stain electron microscopy

AprA ΔACP was prepared following a negative staining protocol[222] and imaged at 120 kV using a Tecnai T12 electron microscope (FEI). A total of 200 micrographs were taken at 52,000x nominal magnification using a 4k by 4k CCD camera (Gatan). EM data processing was performed by Min Su. Images were binned by 2 resulting in a sampling of 4.32 Å per pixel for particle picking and processing. ~50,000 particles were selected automatically, then subject to reference-free 2D classifications using RELION[223], resulting in ~15,000 particles corresponding to the "linear" state and ~15,000 particles corresponding to the "bent" state. The remaining particles were found in classes that corresponded to intermediate states or were in ambiguous orientations. Well-defined class average images in the bent state were selected to generate an initial model using e2initialmodel.py[224], and ~4,000 particles were extracted from the selected classes for 3D reconstruction. The initial model was filtered to 60-Å resolution, and then subjected to 3D auto-refinement with C2 symmetry enforced. The resulting model showed a resolution of ~26 Å (gold standard FSC criterial of 0.143). Crystal structures of the MT_L-ΨGNAT monomer and $MT2_L$ dimer were docked as rigid bodies into the EM density in Chimera[225], as an extensive interface exists between MT_L and ΨGNAT, whereas few contacts are observed between ΨGNAT and $MT2_L$. Additionally, MT_L is not soluble without ΨGNAT, while $MT2_L$ can be independently excised from AprA. Figures of the EM model were prepared using Chimera[225].

Results and Discussion

Origin of the t-butyl group: SAM-derived methyl groups incorporated into apratoxin A

In order to examine the incorporation of methyl groups from SAM into the *t*-butyl group of apratoxin A *in vivo*, [methyl-^{13}C]methionine, the immediate metabolic precursor and methyl donor to SAM, was provided to *M. bouillonii* PNG5-198 cultures. Apratoxin A was subsequently purified from these cultures, and incorporation of ^{13}C into specific carbon atoms in the molecule was monitored by ^{13}C NMR and compared with purified apratoxin A grown without label. Carbons 41-43, comprising the *t*-butyl group in apratoxin A, were enriched by 99.0%, or 1.99-fold in the sample supplemented with [methyl-^{13}C]methionine relative to an unlabeled sample; ^{13}C NMR signals not deriving from the SAM methyl group were essentially unchanged (Figure 5.1a, Table 5.1.) Additionally, in order to support the idea that methylmalonyl-CoA does not provide one of

the methyl groups in the *t*-butyl moiety, we provided the producing culture with [1-^{13}C]propionate, a precursor to methylmalonyl-CoA (MeMal-CoA) in bacterial PKS biosynthesis. We observed only natural-abundance incorporation at C39, the predicted position of possible incorporation into apratoxin A (Figure 5.2, Table 5.2, Table 5.3), indicating that Mal-ACP is the preferred AprA substrate. This is consistent with our previous determination that Mal-ACP is the substrate for AprA MT_L, which performs two methylations to produce Me_2Mal-ACP (Figure 5.1c)[119].

AprA $MT2_L$ is a dual function decarboxylase and methyltransferase

To test the activity of AprA $MT2_L$, we incubated it with the Me_2Mal-ACP product of AprA MT_L in the presence of methyl donor SAM, and analyzed the reaction mixture by intact protein mass spectrometry. Unexpectedly, we observed the production of pivaloyl-ACP, indicating that $MT2_L$ is a bifunctional enzyme, carrying out both decarboxylation and methylation (Figure 5.3a, Figure 5.4a). To verify the generality of this bifunctional activity, we produced the BryX $MT2_L$ domain (40% sequence identity to AprA $MT2_L$) from a homologous module[119] in the bryostatin biosynthetic pathway[20]. BryX $MT2_L$ also produced pivaloyl-ACP (Figure 5.3a, Figure 5.4b), suggesting that BryX is the *t*-butyl-producing module in bryostatin biosynthesis.

Precedent for dual decarboxylation/methyl transfer activity exists in enzymes that function in epigenetic regulation (DNA cytosine-C5 (C5)-methyltransferase)[226, 227] and corrin biosynthesis (CbiT and CobL)[228, 229]. Like AprA $MT2_L$, CbiT and CobL catalyze coordinated decarboxylation and methylation reactions, in which decarboxylation activates a distal carbon for methyl transfer[229], whereas DNA C5-methyltransferase performs decarboxylation independent of methyltransfer[226]. AprA $MT2_L$, DNA C5-methyltransferase, CobL and CbiT are members of the highly divergent class I methyltransferase superfamily, whose members are identified by six conserved motifs in the nucleotide binding core[125]. AprA $MT2_L$ is a distant relative without significant sequence identity to DNA C5-methyltransferase, CbiT or CobL beyond the conserved motifs.

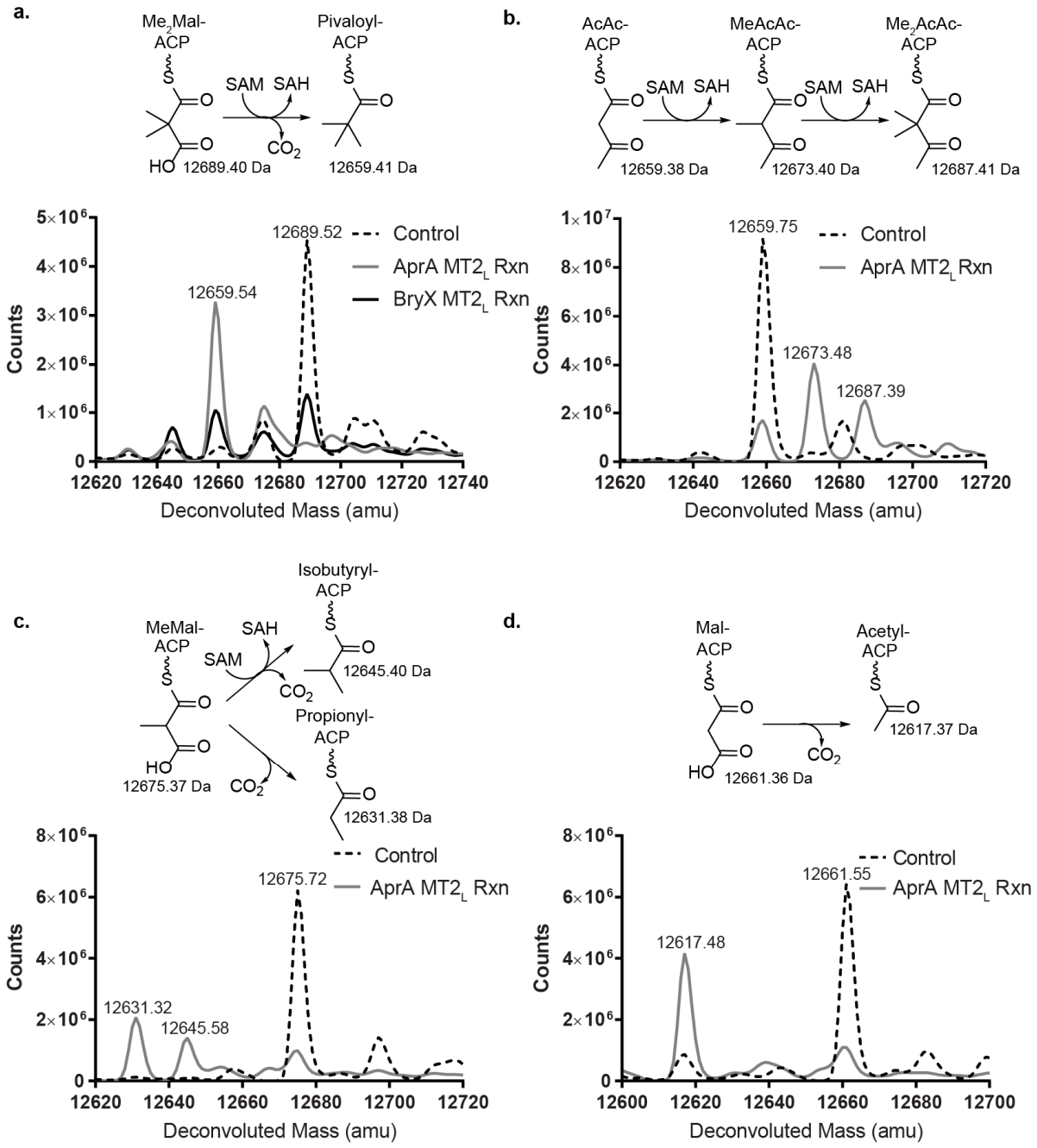


Figure 5.3 Catalytic activity of MT2L

Intact protein mass spectra are shown for AprA ACP species. **a.** AprA MT2L and BryX MT2L reactions on Me2Mal-ACP. AprA MT2L reactions on **b.** AcAc-ACP, **c.** MeMal-ACP, **d.** Mal-ACP. Calculated masses are indicated by the chemical structures, and observed masses on the mass spectra. Negative controls (no enzyme) are indicated in dotted lines.

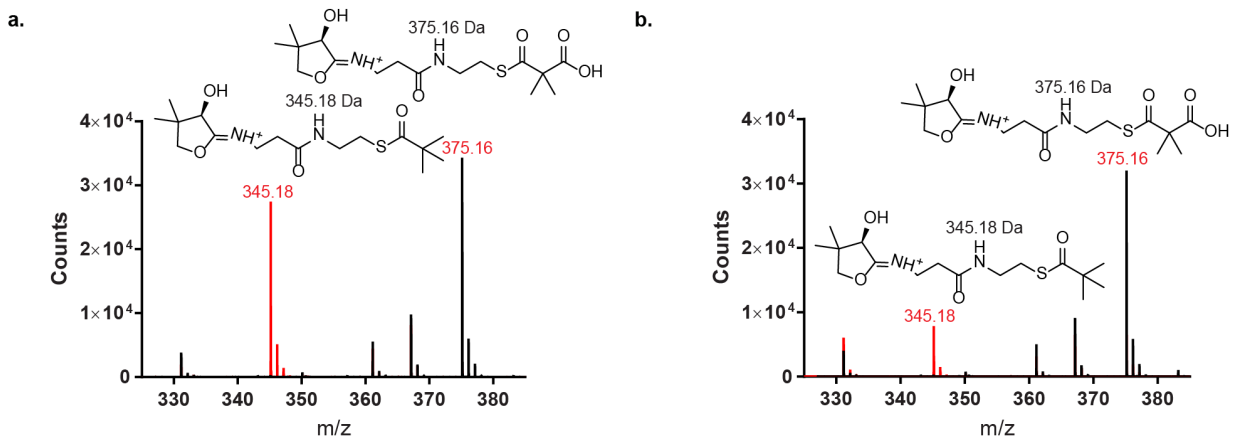


Figure 5.4 Representative mass spectra for the production of Me_2Mal-ACP
Ppant ejection data for Me_2Mal-ACP (generated with AprA MT_L) production by **a.** AprA $MT2_L$ and **b.** BryX $MT2_L$ in Figure 5.3a. Mass spectra from reactions are in red, no-enzyme negative controls in black. Experimentally observed masses are in red, calculated masses in black.

Crystal structure of AprA ΨGNAT-$MT2_L$

To understand how AprA $MT2_L$ mediates its remarkable dual activities, we solved a 2.25-Å crystal structure of the ΨGNAT-$MT2_L$ didomain in complex with *S*-adenosylhomocysteine (SAH) (Figure 5.5, Figure 5.6a, Table 5.5). The AprA ΨGNAT-$MT2_L$ didomain is dimeric in crystals and in solution, as determined by size exclusion chromatography. In the crystals, the head-to-tail dimer interface includes a disulfide bond between the Cys651 residues of the $MT2_L$ domains. A C651S amino acid substitution also yielded dimeric protein, indicating that the ΨGNAT- $MT2_L$ dimer is not a disulfide-dependent artifact of the aerobic lab environment.

The ΨGNAT is truncated relative to functional GNAT homologs[31, 219] and is connected to $MT2_L$ by a 12 residue linker, validating our previous report that the ΨGNAT (truncated GNAT) within the AprA MT_L-ΨGNAT structure is the full domain[119]. AprA $MT2_L$ is similar to the cyanobacterial *C*-MT[117] from the CurJ extension module of curacin A biosynthesis (33% sequence identity, RMSD of 1.3 Å for 273 total Cα atoms, Figure 5.7) and to the fungal citrinin PKS *C*-MT[165] (22% sequence identity, RMSD of 2.1 Å for 261 total Cα atoms), including an extremely long N-terminal helix, a helical lid, and a SAM-binding core with a helical insertion between β-strands 5 and 6. The SAM-binding core of AprA $MT2_L$ and CurJ *C*-MT are virtually identical (RMSD 0.48 Å for 103 core Cα atoms), but their respective lid domains are positioned slightly differently (Figure 5.6b).

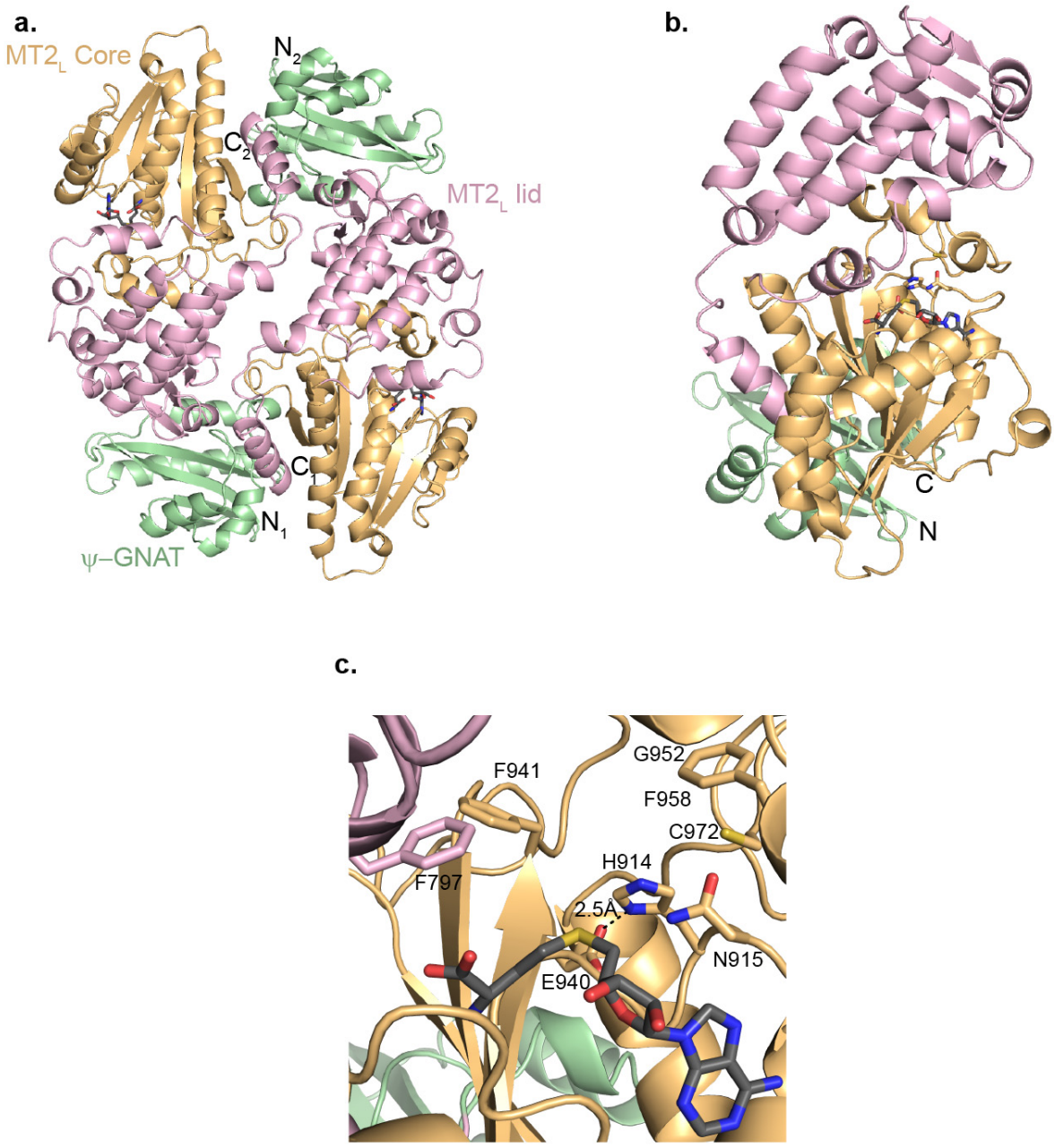


Figure 5.5 AprA ΨGNAT- MT2$_L$ structure and active site
a. AprA ΨGNAT-MT2$_L$ head-to-tail dimer colored by structural region (ΨGNAT, green; MT2$_L$ lid, pink; MT2$_L$ core, orange). SAH is shown in stick form with atomic coloring (C, black; O, red; N, blue; S, yellow). N- and C-termini are labeled for each monomer. **b.** AprA ΨGNAT-MT2$_L$ monomer colored by structural region as in a. **c.** AprA MT2$_L$ active site colored as in a. Amino acids subjected to mutagenesis are shown in sticks.

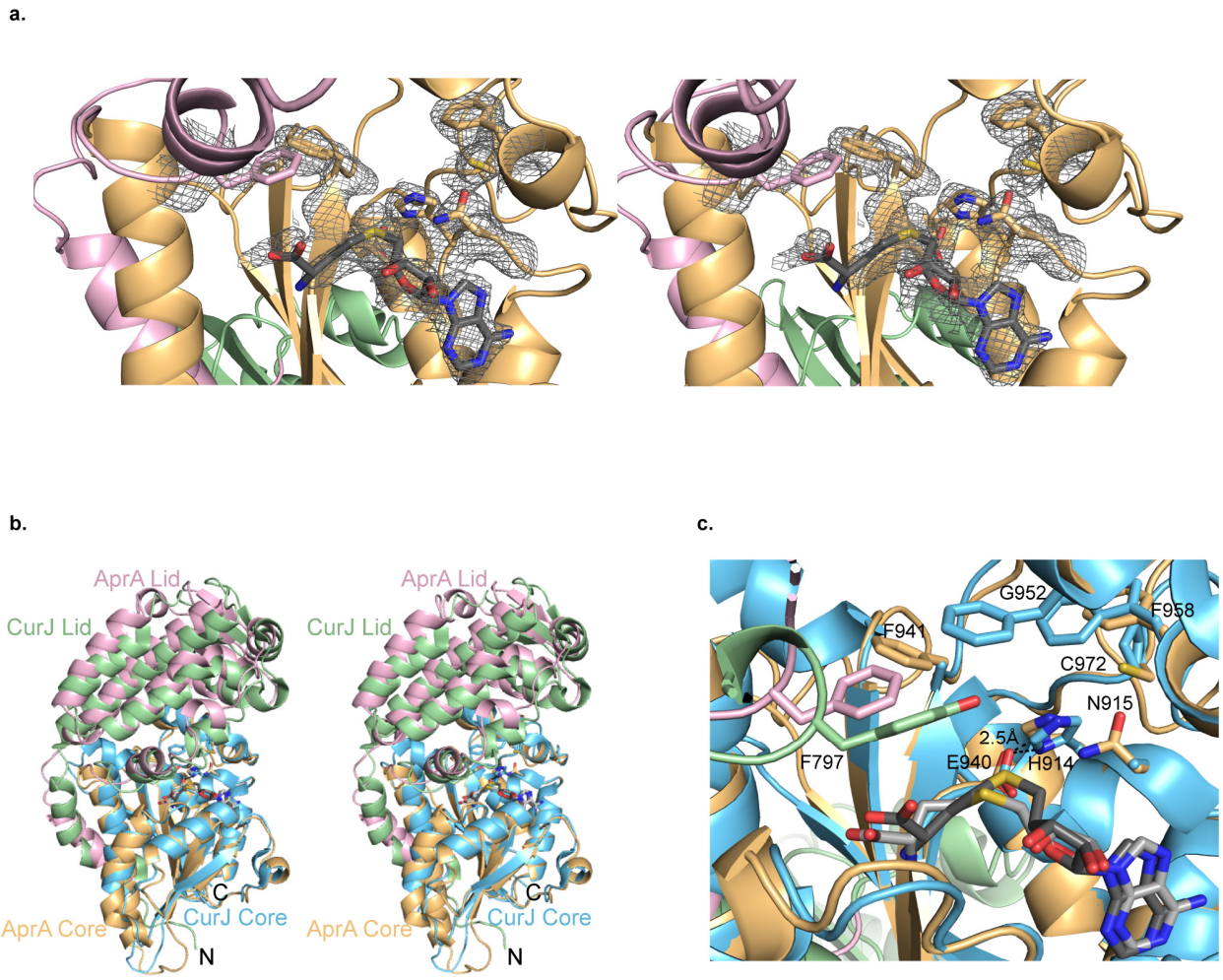


Figure 5.6 AprA MT2$_L$ active site omit density and comparison between AprA MT2$_L$ and CurJ *C*-MT

a. $2F_o$-F_c composite omit density around MT2$_L$ active site contoured at 1σ in stereo. Maps were calculated with simulated annealing (starting temperature 5000 K). **b.** AprA MT2$_L$ (pink lid, orange core) and CurJ *C*-MT (green lid, blue core) superimposed by their core domains (RMSD of 0.48 Å for 103 Cα atoms). SAH is shown in sticks with atomic coloring (C, gray). Key active site residues are shown in sticks. **c.** Zoom view of the AprA MT2$_L$ and CurJ *C*-MT active sites from the superposition in b. Labels correspond to mutagenized amino acids in AprA MT2$_L$.

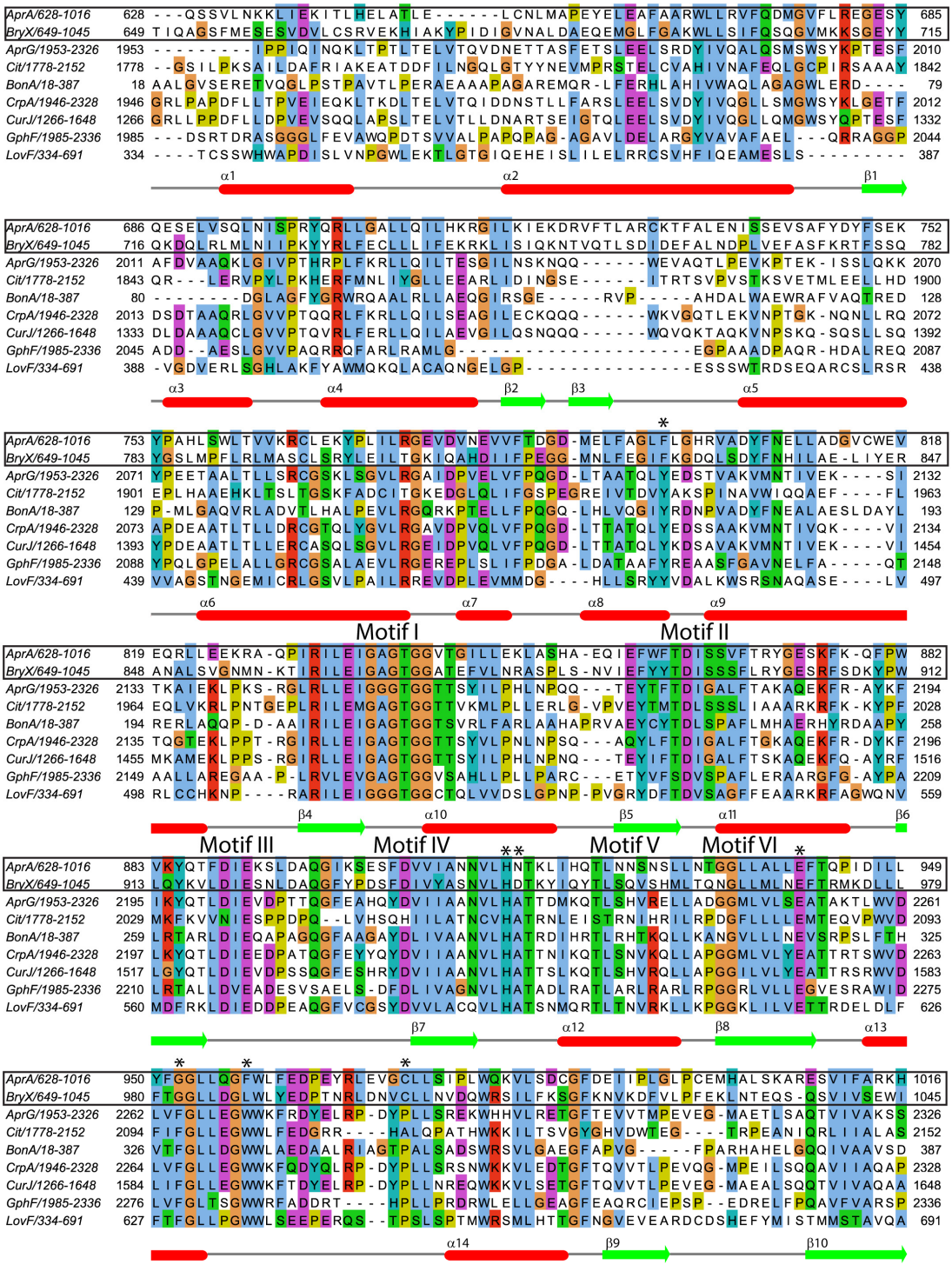


Figure 5.7 AprA MT2$_L$ sequence alignment and secondary structure with BryX MT2$_L$ and PKS *C*-MTs.

AprA amino acids subjected to mutagenesis are starred and class I MT motifs are labeled. Abbreviations (GenBank accession codes): Apr, apratoxin A (WP_075900460); Bry, bryostatin (ABK51302.1); Cit, citrinin (A0A161CEU9.1); Crp, cryptophycin (ABM21569.1); Cur, curacin A (WP_008191795.1); Gph, gephyronic acid (AHA38199.1); Lov, lovastatin (Q9Y7D5.1).

Active site architecture

PKS *C*-MTs, such as CurJ *C*-MT[117] and those in modules AprG and AprI in the apratoxin A pathway[25], catalyze α-methylation of the respective β-keto intermediates, the product of the ketosynthase (KS) extension reaction, but do not act on the carboxylate substrate Mal-ACP [102, 103, 117]. To gain insight into the bifunctional activity of AprA MT2$_L$, we compared the active sites of AprA MT2$_L$ and CurJ *C*-MT[117] (Figure 5.5c, Figure 5.6c), both of which are produced by *Moorea* species. In both structures, SAH sits in a cleft between the lid and core domains with strong electron density for the adenosine moiety and rather weak density for the homocysteine portion (Figure 5.6a), indicating that the bound SAH may not be captured in a catalytic conformation. The cleft between the lid and the core in AprA MT2$_L$ is much shallower than in CurJ *C*-MT, due to the bulky Phe941 and Trp759 side chains, correlating with the shorter Me$_2$Mal-ACP substrate of AprA MT2$_L$ *vs*. the extended β-keto intermediate methylated by CurJ *C*-MT[117]. The His-Glu dyad that is critical for methylation by PKS *C*-MTs[117, 165] is conserved in MT2$_L$ (His914, Glu940) (Figure 5.7). However, several nearby amino acids differ in AprA MT2$_L$ and CurJ *C*-MT. Near the SAH, Phe797 replaces a conserved Tyr that was hypothesized to facilitate methylation by the CurJ *C*-MT through interactions with the SAM sulfonium (Figure 5.6c). Most interestingly, the catalytic His resides in a nearly invariant His-Ala-Thr (HAT) motif in PKS extension-module *C*-MTs (Figures 5.6c, Figure 5.7), while in AprA MT2$_L$, the analogous sequence at the catalytic His914 is His-Asn-Thr.

Other differences exist in a small pocket behind the MT2$_L$ His-Glu dyad. In AprA MT2$_L$, Gly952, Phe958 and Cys972 replace, respectively, Phe, Trp and Pro side chains, which are conserved in extension module *C*-MTs (Figure 5.7). Based on their sequences (Figure 5.7), identical pockets exist in AprA MT2$_L$ and BryX MT2$_L$, which also carries out dual decarboxylation and methylation.

AprA MT2$_L$ is active on multiple substrates

Given the similarity of the AprA MT2$_L$ and CurJ *C*-MT active sites (Figure 5.6c), we tested MT2$_L$ activity on acetoacetyl-ACP (AcAc-ACP), a substrate mimic for CurJ *C*-MT[117]. Similar to the PKS *C*-MT, AprA MT2$_L$ methylated AcAc-ACP to yield dimethyl-acetoacetyl-ACP (Figure 5.3b). Dimethylation by PKS *C*-MTs has been reported previously[100, 101, 103], although no sequence motifs are apparent that delineate the ability to catalyze one or two methylation reactions.

Given the promiscuity of AprA MT2$_L$ for both carboxylated and non-carboxylated acyl groups, we tested MT2$_L$ activity on other potential substrates. Coordinated decarboxylation and methylation activity also occurred with a MeMal-ACP substrate to produce isobutyryl-ACP (Figure 5.3c). Interestingly, this corresponds with the natural occurrence of an isobutyryl group in apratoxin C[230]. However, product formation was approximately 4-fold slower with MeMal-ACP than with Me$_2$Mal-ACP, and the reaction with MeMal-ACP also produced significant quantities of the propionyl-ACP product of decarboxylation only. The slower turnover and shunt product indicate that MT2$_L$ is selective for Me$_2$Mal-ACP (Figure 5.3c). We also tested Mal-ACP as an AprA MT2$_L$ substrate, but, unlike the MT2$_L$ activity with Me$_2$Mal-ACP or MeMal-ACP, we detected only the decarboxylation product acetyl-ACP (Figure 5.3d). Mal-ACP decarboxylation occurred approximately 50-fold slower than the consumption of Me$_2$Mal-ACP, further demonstrating MT2$_L$ selectivity for the dimethylated acyl group. Interestingly, Me$_2$Mal-, MeMal-, and Mal-ACP decarboxylation required SAM; neither SAH nor sinefungin, a stable SAM analog, supported decarboxylation activity. However, SAM was not consumed during the decarboxylation reaction (see below; and Figure 5.8a, b).

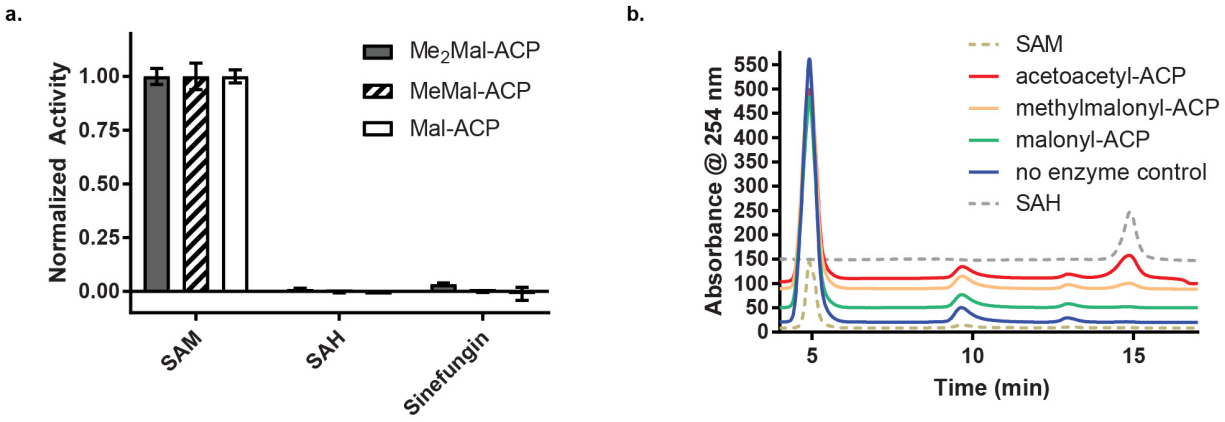


Figure 5.8 AprA MT2$_L$ cofactor dependence
a. Relative activities for reactions of AprA MT2$_L$ with Me$_2$Mal-ACP, MeMal-ACP, Mal-ACP, in the presence of SAM, SAH, and sinefungin. Data represent total product formed. All reaction species were quantified using the Ppant ejection assay[150, 169]. Error bars, in some cases too small to be visible, represent triplicate experiments. **b.** HPLC analysis of SAM consumption in reactions of AprA MT2$_L$ with Mal-ACP, MeMal-ACP and AcAc-ACP. SAM and SAH standards are shown in dashed lines.

Reaction mechanism for AprA MT2$_L$

MT2$_L$-catalyzed methyl transfer was tightly coupled to decarboxylation, as the enzyme had no activity with the decarboxylated intermediates isobutyryl- and propionyl-ACP. The preferred and presumed native substrate was Me$_2$Mal-ACP, which was rapidly and exclusively converted to pivaloyl-ACP (Figure 5.3a). In contrast, reactions on non-native substrates uncoupled decarboxylation and methylation. Substrate MeMal-ACP yielded significant quantities of the propionyl-ACP shunt product of decarboxylation only (Figure 5.3c), and substrate Mal-ACP yielded only the decarboxylated product acetyl-ACP at a significantly slower rate (Figure 5.3d). Coupling of decarboxylation and methylation requires that the reactions occur in one active site. We used site directed mutagenesis to disrupt the conserved His-Glu dyad that is responsible for methylation in PKS *C*-MTs[117, 165], which are AprA MT2$_L$ homologs. Product formation was monitored using the MS-based phosphopantetheine (Ppant) ejection assay, where the Ppant fragment is dissociated from the ACP phosphoserine linkage during ionization and used for quantification based on relative abundance of Ppant species[150, 169]. Activity was evaluated with substrates for coupled decarboxylation and methylation (Me$_2$Mal-ACP and MeMal-ACP), decarboxylation only (Mal-ACP), and methylation only (AcAc-ACP) (Figure 5.9a, 4b, 4c, Figure 5.10).

The His-Glu dyad is essential for both decarboxylation and methylation by MT2$_L$. The H914N and E940A variants had no activity on any substrate, including Mal-ACP, which undergoes decarboxylation only, demonstrating that both the decarboxylation and methylation reactions occur in one active site (Figure 5.5c, Figure 5.9a). We hypothesize that during decarboxylation, His914 accepts a proton and promotes decarboxylation by stabilizing the developing negative charge at the thioester and the resulting enolate (Figure 5.9d). Methyl donation from SAM is likely simultaneous with collapse of the enolate to form pivaloyl-ACP (Figure 5.9d).

As noted above, product formation was abrogated in the absence of SAM, which is consistent with the absolute requirement for this cofactor to initiate decarboxylation, although it is not consumed in the reaction (Figure 5.8a, b). Affinity measurements for substrates in the presence of SAM, SAH, and sinefungin in other methyltransferases[148, 231] indicated that interactions between the active site and the sulfonium, which do not exist in our SAH-bound ΨGNAT- MT2$_L$ crystal structure, may allow this class of enzyme to organize the active site for productive catalysis

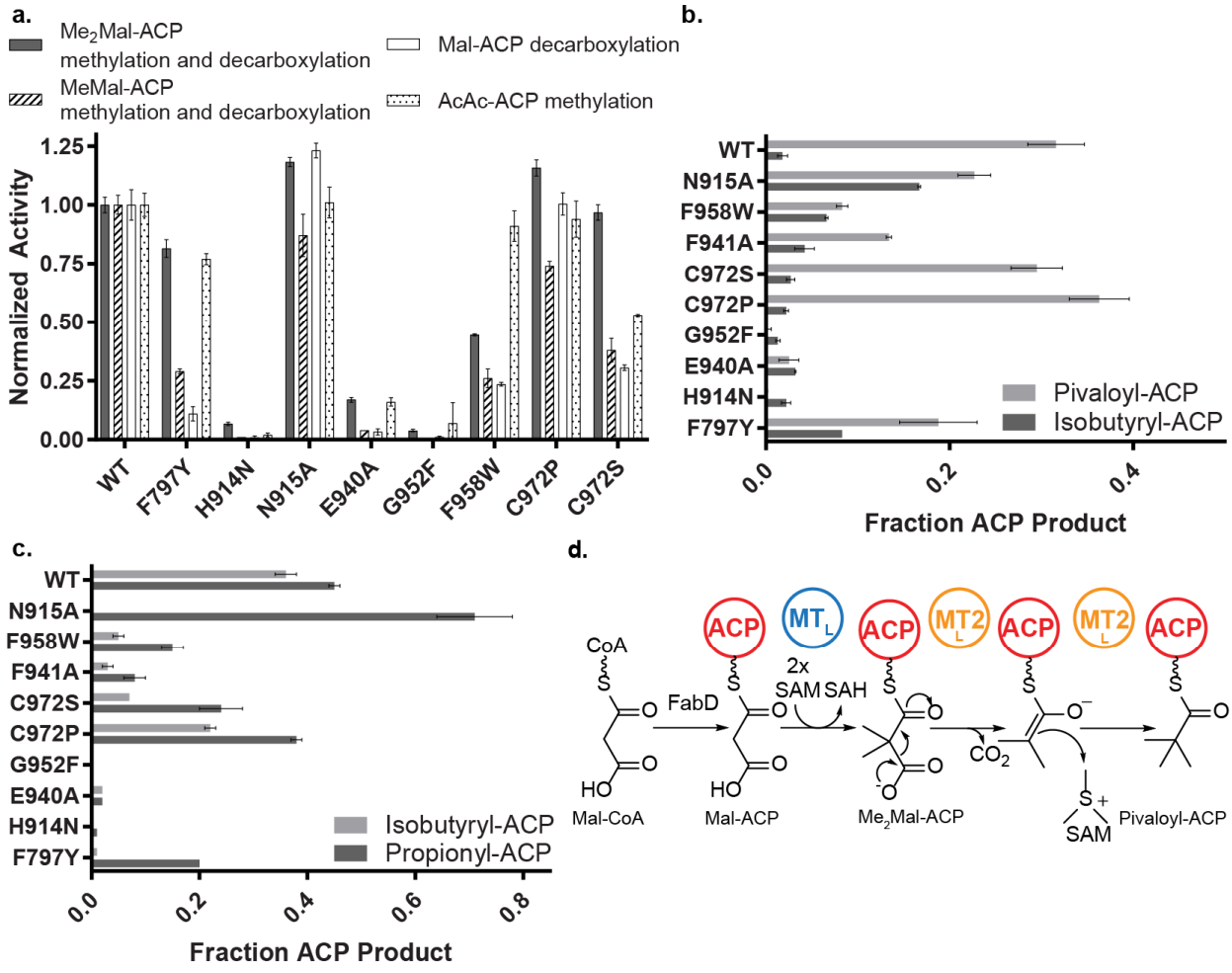


Figure 5.9 Probing AprA MT2$_L$ activity via site-directed mutagenesis
a. Relative activities for reactions of AprA MT2$_L$ with Me$_2$Mal-ACP, MeMal-ACP, Mal-ACP, and AcAc-ACP. Data represent total product formed. Product ratios for data in a. from coupled (decarboxylation + methylation) and uncoupled (decarboxylation only) reactions with **b.** Me$_2$Mal-ACP (pivaloyl-ACP and isobutyryl-ACP), **c.** MeMal-ACP (isobutyryl-ACP and propionyl-ACP). All reaction species were quantified using the Ppant ejection assay[150, 169]. Error bars, in some cases too small to be visible, represent triplicate experiments. **d.** Proposed mechanism for the coupled methylation and decarboxylation sequence to convert Mal-CoA into the pivaloyl-ACP starting unit of apratoxin A biosynthesis

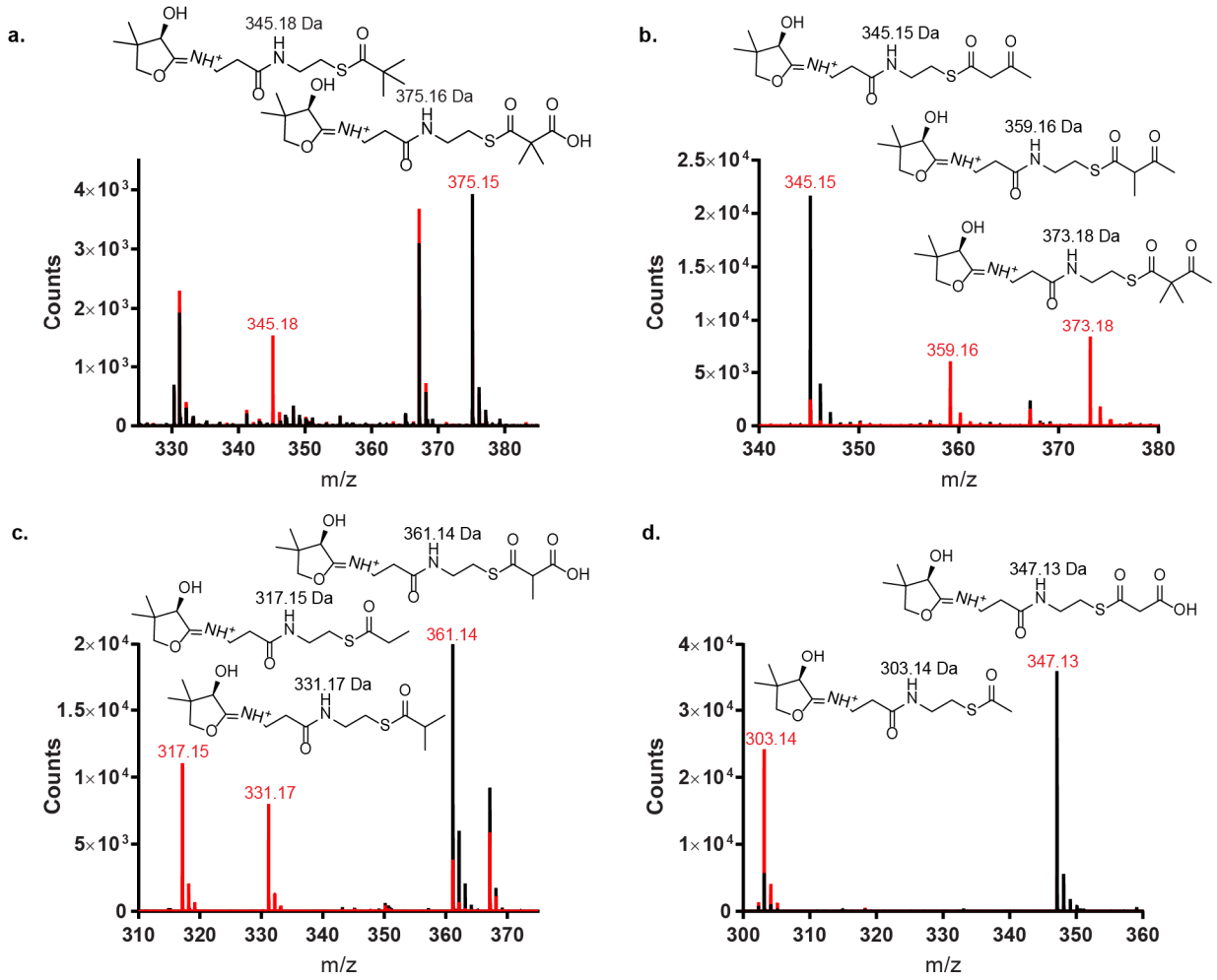


Figure 5.10 Representative mass spectra for coupled decarboxylation and methylation reactions

Representative Ppant ejection data for (**a**) Me_2Mal-ACP (generated with Me_2Mal-thiophenol) (**b**) AcAc-ACP, (**c**) MeMal-ACP and (**d**) Mal-ACP. Reactions with AprA $MT2_L$ are in red, no enzyme control in black. Experimentally observed masses are in red, calculated masses in black.

only when SAM is present by "sensing" small differences that occur upon methyl donor binding. Alternatively, electronic effects of the SAM sulfonium ion, which are not present in SAH or sinefungin, may facilitate the decarboxylation reaction.

In order to understand AprA $MT2_L$ decarboxylation compared to its non-decarboxylating PKS C-MT homologs, such as CurJ *C*-MT, we introduced additional "CurJ like" amino acid substitutions into the AprA $MT2_L$ active site. When Asn915, the amino acid following the catalytic His914, is substituted with Ala, which is conserved in the extension-module *C*-MT HAT motif, we observed near wild type or enhanced turnover (Figure 5.5c, Figure 5.9a, Figure 5.7). However, the N915A substitution decouples decarboxylation and methylation on Me_2Mal-ACP (Figure 5.9b)

and MeMal-ACP (Figure 5.9c), as indicated by the increase in production of the respective isobutyryl-ACP and propionyl-ACP shunt products resulting from decarboxylation only.

Decoupling of decarboxylation and methylation of Me_2Mal-ACP and MeMal-ACP also occurred when $MT2_L$ Phe797 was substituted with Tyr (Figure 5.9b, c), which is conserved in nearly all extension-module *C*-MTs (Figure 5.6c, Figure 5.7). Notably, the F797Y variant nearly eliminated decarboxylation activity on Mal-ACP, but had little effect on methylation of AcAc-ACP (Figure 5.9a). We next examined the small pocket behind the His-Glu dyad, which is unique to the bifunctional AprA $MT2_L$ and BryX $MT2_L$. An F958W substitution also decoupled decarboxylation and methylation of Me_2Mal-ACP (Figure 5.9b) and MeMal-ACP (Figure 5.9c) and, like the F797Y substitution, had little effect on methylation of AcAc-ACP (Figure 5.9a). Occluding the small pocket with a G952F substitution eliminated all activities, whereas substitutions at Cys972 within the pocket (C972P, C972S) had modest effects.

As small amino acid substitutions to the AprA $MT2_L$ active site (N915A, F979Y, F958W) and minor substrate alterations (*e.g*. MeMal-ACP *vs*. Me_2Mal-ACP) decoupled decarboxylation and methylation, we conclude that AprA $MT2_L$ evolved from a methylation-only enzyme to promote formation of pivaloyl-ACP. Given the weak electron density for SAH in the AprA $MT2_L$ crystal structure (Figure 5.6a), several amino acids are likely not in their final catalytic positions. Therefore, we hypothesize that once in catalytically competent positions, amino acids that couple methylation and decarboxylation (Asn915, Phe797, Phe958 and perhaps others) either position the carboxylated substrates for catalysis or stabilize the proposed enolate intermediate (Figure 5.9d). Stabilization of the enolate and substrate positioning adjacent to the SAM methyl donor are essential as the decarboxylated intermediate could readily collapse to the shunt product by accepting a proton without methyl transfer.

FabD provides the initial acyl transfer step

As neither AprA MT_L-ΨGNAT nor ΨGNAT performed the malonyl transfer reaction (Mal-CoA to Mal-ACP) to initiate apratoxin A biosynthesis[119], we tested AprA $MT2_L$ for malonyl acyltransferase activity, but detected no transfer of malonyl from CoA to AprA ACP (Figure 5.11a). Additionally, the *apr* gene cluster encodes no other candidate acyltransferase for the initiation reaction.

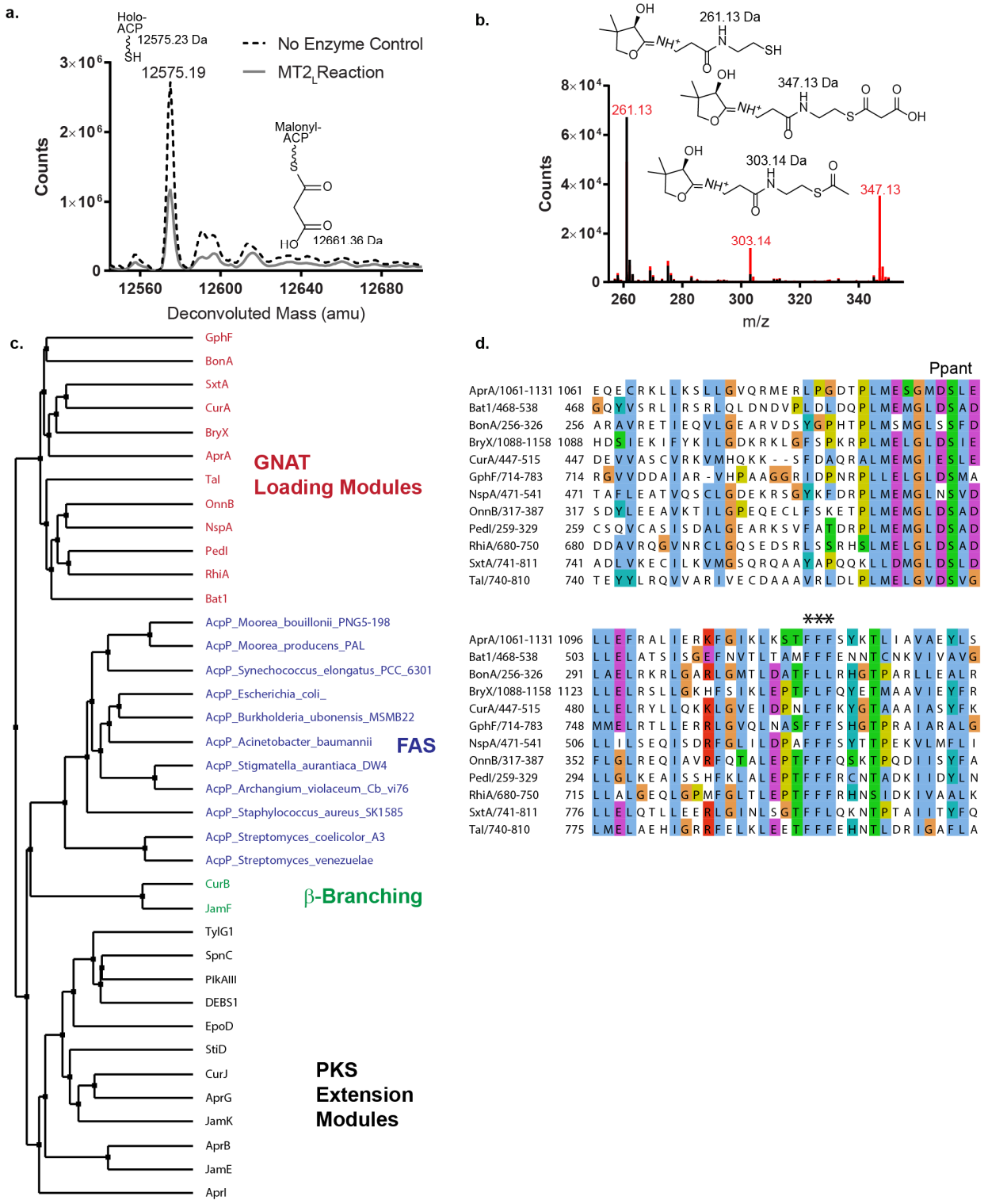


Figure 5.11 Acyltransfer activity and GNAT initiation-module-associated ACPs
a. Intact protein mass spectra of AprA ACP loading reactions of MT2L with Mal-CoA and holo-ACP. Calculated masses are indicated next to chemical structures, observed masses are indicated on spectra. No acyltransfer was observed in a 1 hr reaction. **b.** Representative Ppant ejection data

for reactions of FabD with Mal-CoA and AprA holo-ACP. Reaction with FabD is in red, no enzyme control in black. The presence of acetyl-ACP is due to in source decay during ionization. **c.** Dendrogram of ACPs from GNAT initiation modules (red), fatty acid synthase AcpP (blue), PKS β-branching (green), PKS extension modules (black). **d.** Sequence alignment of GNAT associated ACPs. Ppant attachment site (Ser1093) is labeled. Phenylalanine-containing motif specific to ACPs in GNAT initiation modules is marked with asterisks. Pathway abbreviations(GenBank accession codes) are as follows: Apr, apratoxin A (WP_075900460); Bry, bryostatin (ABK51302.1); Cur, curacin A (AEE88289.1); Sxt, saxitoxin (WP_009343302.1); Ta, myxovirescin A (WP_011553948.1); Gph, gephyronic acid (KF479198.1); Rhi, rhizoxin (WP_013435483.1); Nsp, nosperin (ADA69237.1); Bat, batumin (WP_052451043.1); Onn, onnamide (AAV97870.1); Ped, pederin (AAR19304.1); Bon, bongkrekic acid (AFN27480.1).

Previous studies showed that the malonyl-acyltransferase (FabD) of fatty acid biosynthesis in the producing organism can compensate for the lack of an acyltransferase in the initiation module of some PKS systems[196, 232]. Therefore, we identified *fabD* in the *M. bouillonii* PNG5-198 genome[184], cloned the gene, and produced recombinant FabD in *E. coli*. Incubation of FabD with AprA holo-ACP led to rapid formation of Mal-ACP (Figure 5.12, Figure 5.11b). Surprisingly, AprA ACP shares only 20% sequence identity with the *M. bouillonii* fatty acid synthase (FAS) ACP (AcpP). Additionally, ACPs associated with GNAT-containing initiation modules, such as AprA, clade separately from both PKS extension-module ACPs and FAS ACPs (Figure 5.11c) and can be distinguished by a phenylalanine motif (Figure 5.11d). Therefore, we tested the specificity of FabD for the AprA initiation-module ACP relative to extension-module ACPs in the apratoxin A pathway using the AprI ACP (23% identity to AprA ACP). AprI is a typical PKS extension module with an embedded acyltransferase. FabD catalyzed malonyl transfer to AprI ACP, but was twofold faster for malonyl transfer to AprA ACP (Figure 5.12). This is consistent with a previous report that FabD can compensate for catalytically inactive acyltransferase domains in a modular PKS[233]. We propose that *M. bouillonii* exploits primary metabolism by using FabD from fatty acid biosynthesis to initiate apratoxin A biosynthesis by providing the malonyl starter unit to the AprA ACP.

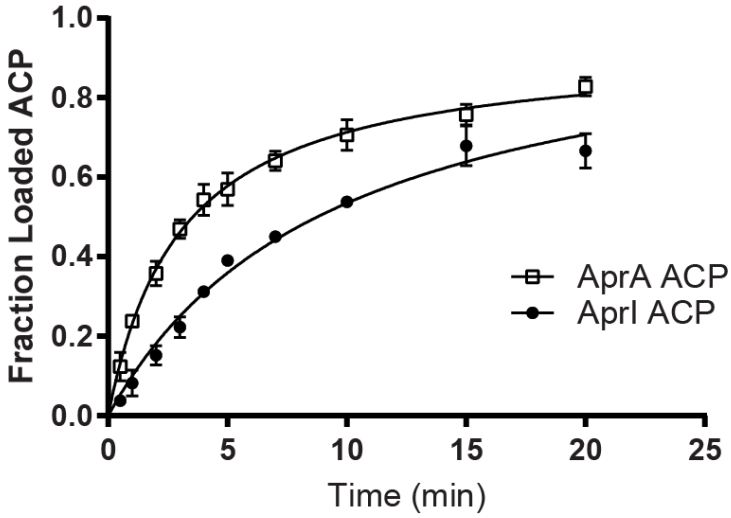


Figure 5.12 FabD malonyl acyltransfer activity
Malonyl is transferred from CoA to ACP by *M. bouillonii* PNG5-198 FabD. Reaction was monitored using the Ppant ejection assay[150, 169]. Error bars represent triplicate experiments and, in some cases, are too small to be visible.

Modeling and EM visualization of the AprA module

The crystal structures of the AprA MT_L-ΨGNAT[119] and ΨGNAT- $MT2_L$ fragments provided a unique opportunity to model the overall architecture of AprA (MT_L-ΨGNAT- $MT2_L$) by superimposing the ΨGNAT domains of the two structures (RMSD of 1.0 Å for 107 Cα). The composite model (Figure 5.13a) lacks steric clashes as MT_L and $MT2_L$ contact opposite faces of the ΨGNAT. The $MT2_L$ dimer sits at the center of the curved AprA model with the MT_L domains on opposite sides. The ACP is missing in our model, but is connected to the C-terminus of the $MT2_L$ dimer via a 44 amino acid linker.

We tested the AprA model using electron microscopy to visualize intact AprA ΔACP (Figure 5.13b, Figure 5.14a, b). Consistent with the crystal structures and the model, the negative-stain 2D class averages revealed a structure with overall twofold-symmetry. A large central lobe, consistent with the $MT2_L$ dimer (96 kDa), is flanked by two smaller lobes, each consistent with monomeric MT_L-ΨGNAT (75 kDa, Figure 5.13b). MT_L is insoluble without ΨGNAT[119] whereas $MT2_L$ is a stable dimer, and accordingly the flanking MT_L-ΨGNAT has few contacts with $MT2_L$, suggesting that the ΨGNAT is associated with MT_L. The overall architecture of the AprA model is similar to classes corresponding to ~30% of total particles, where the two MT_L-ΨGNAT lobes

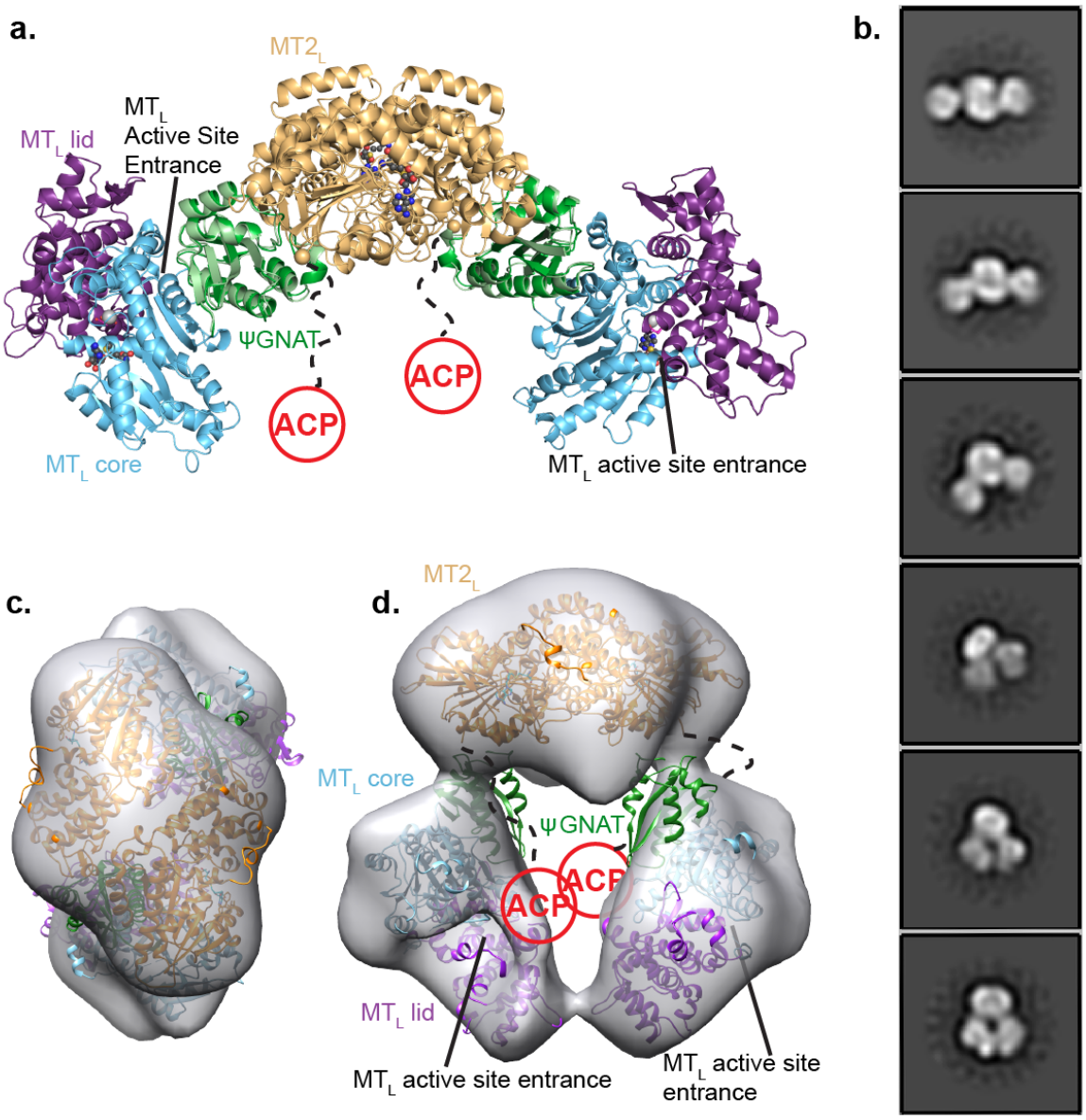


Figure 5.13 Architecture of the AprA module

a. AprA model based on crystal structures of MTL-ΨGNAT and ΨGNAT-MT2L fragments, colored by structural region (MTL lid, purple; MTL core, blue; ΨGNAT, green; MT2L dimer, orange). SAM and SAH are shown in spheres with atomic coloring (gray C). In the MTL active site, Mn^{2+} is shown in light gray spheres and malonate in sticks (pink C). The model was created by superposition of the ΨGNAT domains of the two crystal structures (RMSD 1.0 for 107 Cα αtoms). The 44 amino acids linking MT2L and ACP are depicted in a dashed line. **b.** Selected negative stain class averages of AprA (ΔACP) showing the central MT2L dimer and ΨGNAT-MT2L wings. Linear and bent states of AprA are observed in these images. **c.** 3D reconstruction of AprA in the bent conformation (gray), representing ~30% of the particles (top down view). Docked crystal structures of AprA MTL-ΨGNAT and AprA MT2L are colored as in a. **d.** side view of 3D reconstruction with docked AprA domains.

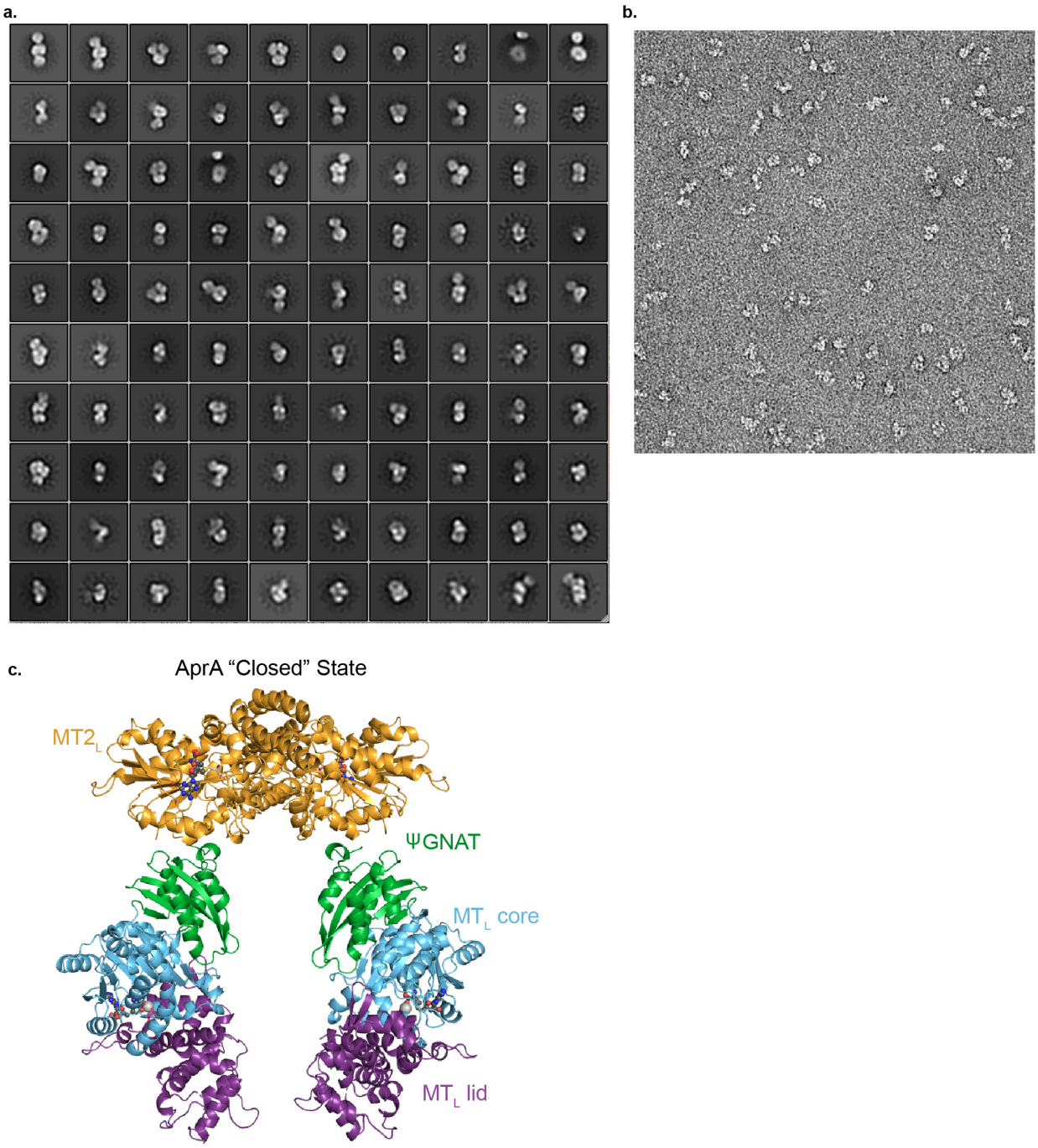


Figure 5.14 EM class averages, micrographs and AprA model

a. Negative stain class averages of AprA ΔACP. **b.** Raw negative stain micrograph of AprA ΔACP. **c.** Model of AprA based on a 3D reconstruction of bent-state particles, colored as in Figure 5.13. Compared to the linear-state model from superposition of the MT_L-ΨGNAT and ΨGNAT-$MT2_L$ crystal structures, the MT_L-ΨGNAT is rotated ~45° relative to the $MT2_L$ dimer.

are linear relative to $MT2_L$. However, a similar number of particles belong to classes with a bent arch shape where the MT_L-ΨGNAT lobes appear repositioned relative to the $MT2_L$ dimer (Figure 5.13b).

The bent-state class averages were more uniform and of higher quality than the linear-state classes. Thus, we extracted particles of the bent state and obtained a 3D reconstruction to test whether the size and shape of the EM density was consistent with the crystal structures (Figure 5.13c, d). The $MT2_L$ dimer crystal structure docks nearly perfectly into the EM density, and the size and shape of the MT_L -ΨGNAT correspond well to the flanking density. In the bent conformation, MT_L is severely rotated relative to its position in the model based on the crystal structures (Figure 5.13a, d and Figure 5.14c). In this position, the MT_L active site entrance is inside the chamber of the AprA arch, ~50 Å from the ACP attachment site.

We hypothesize that AprA adopts different conformations during its catalytic cycle, and that this may protect the substrate from aberrant reactions. In one conformation, the ACP must be accessible to FabD to receive the malonyl starter unit. Given the promiscuity of AprA $MT2_L$, Mal-ACP must be protected from the $MT2_L$ active site to prevent premature decarboxylation prior to MT_L methylation. The MT_L reaction may occur in the bent state where the MT_L active site entrance is closer to the ACP attachment site. Once MT_L has formed Me_2Mal-ACP, the MT_L-ΨGNAT may swing to an open conformation to allow access to the $MT2_L$ active site, where the decarboxylative methylation forms pivaloyl-ACP.

In conclusion, our complete characterization of the apratoxin A initiation module AprA describes a unique biochemical process to form *t*-butyl groups in natural product biosynthesis (Figure 5.9d) and clarifies a previously nebulous route to an unusual chemical functionality. First, the FAS malonyl-acyltransferase FabD loads the AprA ACP with malonyl from CoA. Malonyl-ACP is then dimethylated by the mononuclear iron-dependent methyltransferase AprA MT_L[119]. Dimethylmalonyl-ACP undergoes a coordinated decarboxylation and methylation reaction by AprA $MT2_L$ to form pivaloyl-ACP. The functional annotation of AprA serves as an identifier for gene clusters producing *t*-butyl-containing natural products, especially from marine sources, as exemplified by our identification and characterization of BryX $MT2_L$ from the bryostatin biosynthetic pathway. Our structural characterization of the full AprA module provides insight into the overall architecture and mobility of this remarkable multifunctional module.

Chapter 6 GNAT-like Domains for Initiation of Polyketide Biosynthesis are Primarily Acyl-ACP Decarboxylases

Summary

(GCN5)-related *N*-acetyltransferase (GNAT) proteins are typically associated with acetylation of a variety of amine substrates from an acetyl-CoA donor. A decarboxylation activity exists in some GNAT superfamily members, including malonyl-CoA decarboxylase (MCD) and enzymes for initiation of the biosynthesis of some polyketide secondary metabolites. Here, we probe the role of GNAT-like domains that are proposed to possess the dual functions of *S*-acyl transfer to an acyl carrier protein (ACP) and decarboxylation in polyketide biosynthesis. The GNAT-like domain within GphF in the gephyronic acid biosynthetic pathway selectively catalyzes decarboxylation of dimethylmalonyl (Me_2Mal-) ACP, but lacks acyl transfer activity. Biochemical characterization and the crystal structure of a GphF GNAT complex with an isobutyryl-CoA product analog provide insight into the mechanism of decarboxylation by GNAT superfamily members. Analysis of CurA GNAT, another polyketide synthase (PKS) GNAT-like domain with proposed dual function, suggests that, like the GphF GNAT, the CurA GNAT catalyzes decarboxylation only. Thus, it is possible that an enzyme outside the biosynthetic pathway catalyzes malonyl acyl transfer. As the primary role of PKS GNAT-like domains is decarboxylation, we propose to reclassify GNATs in PKS initiation modules as acyl-ACP decarboxylases.

Introduction

Named for the founding member, general control non-repressible 5 (GCN5)-related *N*-acetyltransferase (GNAT)[234], the GNAT superfamily of enzymes spans all kingdoms of life and is best known for using an acetyl-coenzyme A (CoA) donor to acetylate amine substrates, such as histones, aminoglycoside antibiotics, arylalkylamines, and a variety of other proteins and metabolic intermediates[42, 43, 235]. Other GNAT superfamily members use alternative acyl donors, such as myristol-CoA[236], aminoacyl-tRNAs[237], or fatty acyl-ACPs[238-240]. Although studied for decades[241], malonyl-CoA (Mal-CoA) decarboxylase (MCD) was identified as a member of the

GNAT superfamily only when a crystal structure was solved[183, 242]. MCD, which has very low sequence identity (~10%) to *N*-acyltransferase members of the GNAT superfamily, is conserved from bacteria to mammals[243, 244] and catalyzes decarboxylation and not acyl transfer. In mammals, MCD modulates levels of malonyl-CoA and in some cases can promote the use of methylmalonyl-CoA as a substrate for fatty acid biosynthesis, leading to tissue-specific production of multimethyl-branched fatty acids[244]. Disruption of the gene encoding MCD results in malonic aciduria[245]. Bacterial genes for MCD are linked to those for a Mal-CoA synthetase and malonate transporter[243]. Given that proteins with the GNAT fold can catalyze both acyl transfer and decarboxylation, the common function of the superfamily may be to bind the phosphopantetheine (Ppant) moiety of CoA or ACP.

Natural product biosynthetic pathways frequently repurpose enzymes found elsewhere in biology for the production of molecules that confer a selective advantage on the producing organism. One such example is the acquisition of a GNAT-like enzyme in the polyketide biosynthetic pathways of a variety of bacteria (Figure 6.1, Figure 6.2) [21, 23, 25, 31, 45, 93-98, 116]. Polyketides are synthesized in a stepwise manner from acyl-CoA building blocks. Substrates and products in the modular multienzyme type I polyketide synthase (PKS) enzyme assembly lines are tethered to the phosphopantetheine (Ppant) cofactor of an ACP. PKS GNAT-like domains were first identified as members of the GNAT superfamily in sequence analysis of the pederin biosynthetic pathway[44]. The GNAT-like domain is part of the "loading" or initiation module, where it is expected to transfer an acetyl-group from acetyl-CoA to the Ppant of the initiation module ACP (ACP_L). Identification of the diphosphate binding P-loop for acetyl-CoA binding in other GNAT family members suggested that the pederin GNAT-like domain shares the canonical acetyl-CoA binding property of the GNAT fold. Additional GNAT-like domains exist in PKS initiation modules that produce acetyl[23, 93, 94, 96, 98, 116], propionyl[45, 97], isobutyryl[21], and *t*-butyl[20, 25, 120] starter units.

The initial characterization of a PKS GNAT-like domain was for the CurA GNAT in the pathway for the antimitotic curacin A. We reported a slow *S*-acetyl transfer activity from acetyl-CoA to the ACP_L for the CurA GNAT (Figure 6.1a)[31]. Unexpectedly, CurA GNAT rapidly decarboxylated Mal-ACP or Mal-CoA prior to acetyl transfer to generate an acetyl starter unit for curacin A biosynthesis[31]. Decarboxylation by CurA GNAT, which required conserved His and Thr amino acids, represented a new catalytic function for the GNAT superfamily. MCD had not yet

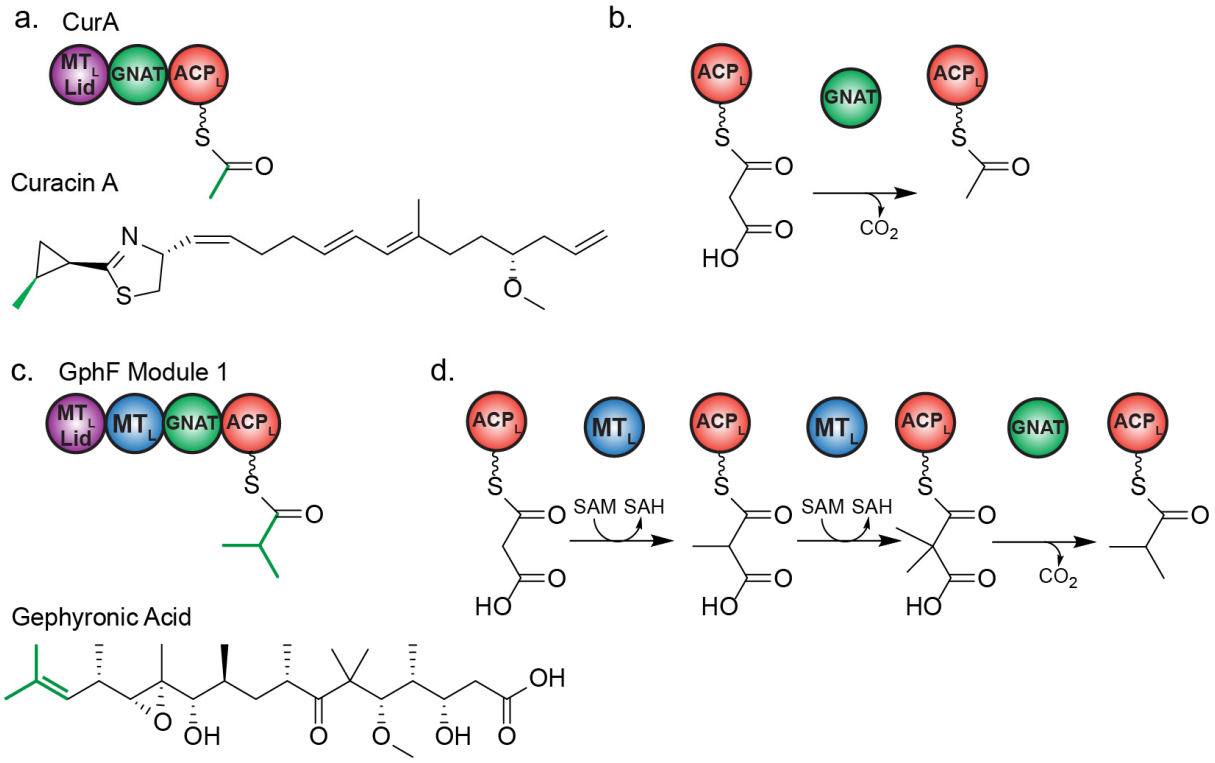


Figure 6.1 GNAT-like enzymes in gephyronic acid and curacin A biosynthesis
a. The CurA initiation module produces an acetyl starter unit (green). The CurA N-terminus is homologous to an MT_L lid, but the SAM binding core domain has been lost. **b.** The CurA GNAT-like enzyme decarboxylates Mal-ACP to acetyl-ACP. **c.** The GphF initiation module produces an isobuturyl starter unit (green). **d.** GphF MT_L dimethylates Mal-ACP, and the GNAT-like domain decarboxylates Me_2Mal-ACP to isobutyryl-ACP.

been identified as a GNAT superfamily member, as it shares only 13% sequence identity with CurA GNAT. Later, and by reference to the CurA GNAT result, the analogous His and Ser residues were implicated in MCD decarboxylation[183]. Like the CurA GNAT, the GNAT-like domain in the saxitoxin PKS initiation module displayed dual activities to CurA GNAT, including a weak acyl transfer activity and a preference for decarboxylation of methylmalonyl (MeMal)-ACP to propionyl-ACP, the expected saxitoxin starter unit[219]. In apratoxin A and bryostatin biosynthesis, the GNAT-like domain is truncated and possesses no detectable decarboxylation or acyl transfer activity[119, 120]. The initial acyl transfer step to commence apratoxin A biosynthesis may require the bacterial fatty acid synthesis enzyme malonyl-acyltransferase (MAT) FabD[120].

The gephyronic acid biosynthetic pathway begins with the GphF initiation module and produces a cytostatic polyketide[21]. The GphF initiation module contains a methyltransferase, a

GNAT-like domain, and an ACP (MT_L-$GNAT_L$-ACP_L, Figure 6.1c) and is predicted to generate an isobutyryl starter unit in contrast to other characterized GNAT-containing initiation modules, which produce acetyl (curacin A)[31], propionyl (saxitoxin)[219], and pivaloyl (apratoxin, bryostatin) starter units[120]. Isotope labeling studies indicated that the gephyronic acid isobuturyl unit is derived from *S*-adenosylmethionine (SAM)[21]. Previously, we showed that MTs associated with PKS initiation modules, including GphF MT_L, are unusual mononuclear-iron-dependent methyltransferases with a large lid domain (formally known as adaptor region) that catalyze methylation of Mal-ACP to MeMal-ACP or dimethylmalonyl (Me_2Mal)-ACP (Figure 6.1d)[119].

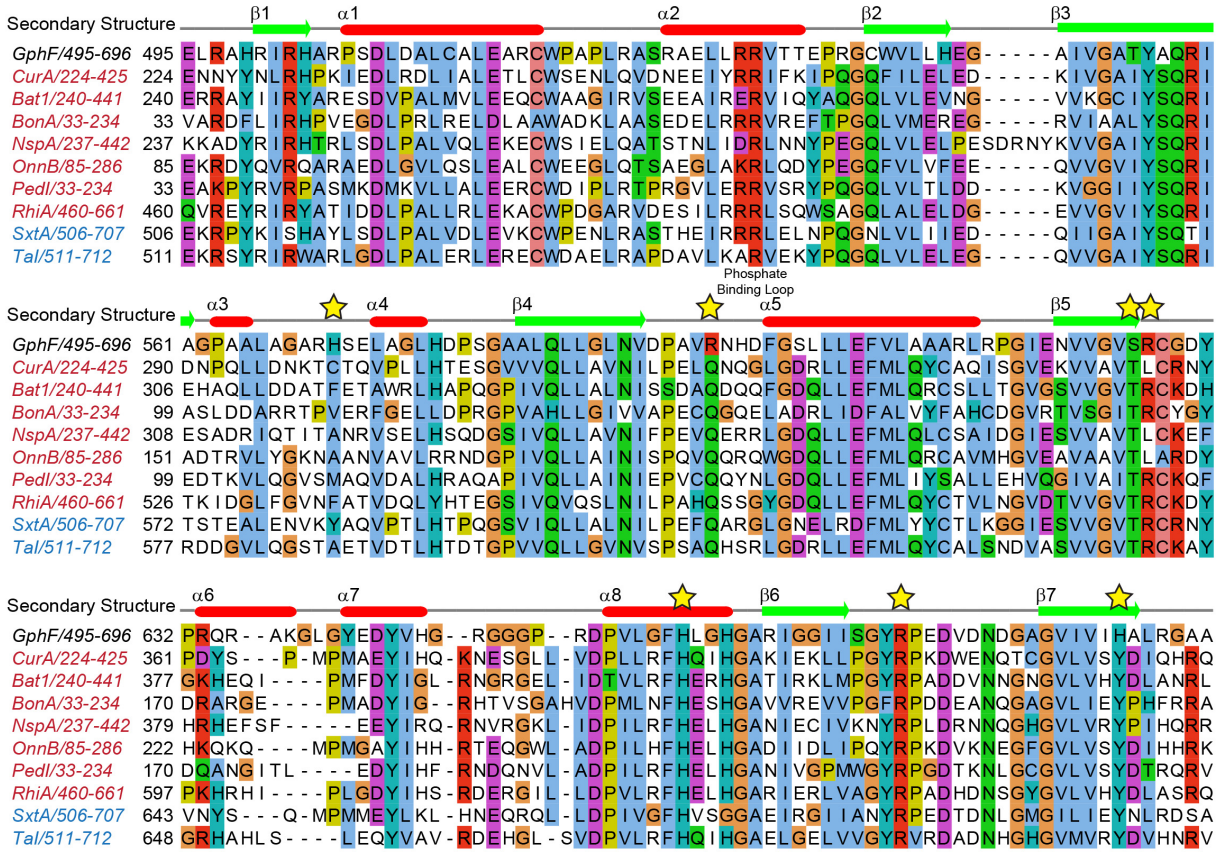


Figure 6.2 Sequence alignment of GNAT-like domains from PKS pathways for natural products of known structure

All GNAT-like domains occur in presumed initiation modules. Only GphF is predicted to produce an isobutyryl starter unit; initiation modules with propionyl starter units are in blue and those with acetyl starter units in red. The GphF GNAT secondary structure annotation is displayed above the alignment. Amino acids subjected to mutagenesis are starred. Protein abbreviations (GenBank accession codes) are as follows: Gph- gephyronic acid (KF479198.1), Cur- curacin A (AEE88289.1), Bat- batumin (WP_052451043.1), Bon- bongkrekic acid (AFN27480.1), Nsp- nosperin (ADA69237.1), Onn- onnamide (AAV97870.1), Ped- pederin (AAR19304.1), Rhi-

rhizoxin (WP_013435483.1), Sxt- saxitoxin (WP_009343302.1), Ta- myxovirescin A (WP_011553948.1).

The carboxylate of Mal-ACP is essential for methyl transfer activity, as no methylation was observed on acetyl-, propionyl-, or isobutyryl-ACP. Decarboxylation of Me_2Mal-ACP by the GphF GNAT-like enzyme would result in the predicted isobutyryl starter unit (Figure 6.1d).

Here, we investigate the catalytic role of the GNAT-like domain in polyketide biosynthesis. Biochemical comparison of the GphF and CurA GNAT-like enzymes reveals a strong acyl-group selectivity for decarboxylation substrates. Additional insight into the decarboxylation reaction is provided by a 2.8-Å crystal structure of the GphF GNAT-like enzyme with the product analog isobutyryl-CoA. Surprisingly, no acyl transfer activity was detected for the GphF GNAT-like domain. Given the low rate of detectable acyl transfer compared to decarboxylation for other PKS GNAT-like enzymes[31, 219], their function appears to be primarily decarboxylation. Thus, we propose to reclassify GNAT-like proteins in PKS pathways as acyl-ACP decarboxylases.

Experimental Procedures

Construct design

All primers are listed in Table 6.1. A cDNA encoding the GphF GNAT was amplified from a partial *gphF* clone provided by Richard Taylor (Notre Dame University) and inserted into pMCSG7 by ligation independent cloning (LIC) to create the expression plasmid for GphF GNAT (residues 498-705, pMAS/APS340). A cDNA encoding the CurA ACP (residues 444-521) was amplified from a cosmid library[23] and inserted into pMCSG7 via LIC to create pMAS/APS442. All *gphF* site-directed mutants (Table 6.1) were introduced into pMAS340 using the QuickChange protocol (Stratagene). All constructs and mutations were verified by Sanger sequencing at the University of Michigan DNA Sequencing Core.

Protein expression and purification

Escherichia coli strain BL21(DE3) was transformed with plasmids encoding GphF GNAT and CurA GNAT (in pMCSG7)[31]. Transformed cells were grown in 0.5 L of TB media at 37°C supplemented with 100 μg mL^{-1} ampicillin to an OD_{600}=1-2, cooled to 20°C for 1 hr, and induced with 200 μM IPTG for 18 hr. Cell pellets were resuspended in 35 mL Tris buffer A (50 mM Tris pH 7.4, 300 mM NaCl, 10% (v/v) glycerol, 20 mM imidazole), 0.1 mg mL^{-1} lysozyme, 0.05 mg

mL^{-1} DNase, and 2 mM $MgCl_2$, incubated on ice for 30 min, lysed by sonication, and cleared by centrifugation (38,760 x g, 15-30 min, 4°C). The supernatant was filtered and loaded onto a 5 mL

Table 6.1 Primers for cloning Gph and Cur GNAT constructs

GphF 498 F	pMAS/ APS 340	**TACTTCCAATCCAATGCA**GCGCATCGGATCCGTC
GphF 705 R		**TTATCCACTTCCAATGTTA**GCTCGCGGGGACC
CurA 444 F	pMAS/ APS 442	**TACTTCCAATCCAATGCA**GAAGATATAGATGAAGTTGTGGCTAG
CurA 521 R		**TTATCCACTTCCAATGTTA**GCTTCTCTCTGTTGTCTCCC
GphF H571C F	pMAS 343	CGGCGCCTCTGCAGGCTTCGCGGGCC
GphF H571C R		GGCCCGCGAAGCCTGCAGAGGCGCCG
GphF R597Q F	pMAS 387	GACCCCGCGGTCCAGAACCACGACTTCG
GphF R597Q R		CGAAGTCGTGGTTCTGGACCGCGGGGTC
GphF S626A F	pMAS 426	TCGTCGGCGTCGCGCGTTGCGGC
GphF S626A R		GCCGCAACGCGCGACGCCGACGA
GphF S626T F	pMAS 388	TCGTCGGCGTCACGCGTTGCGGC
GphF S626T R		GCCGCAACGCGTGACGCCGACGA
GphF H660A F	pMAS 427	GGTGCTCGGGTTCGCCCTCGGTCACGGC
GphF H660A R		GCCGTGACCGAGGGCGAACCCGAGCACC
GphF H660N F	pMAS 438	GGTGCTCGGGTTCAACCTCGGTCACGG
GphF H660N R		CCGTGACCGAGGTTGAACCCGAGCACC
GphF R675K F	pMAS 419	GATCATCAGCGGCTACAAGCCCGAGGATGTCGAC
GphF R675K R		GTCGACATCCTCGGGCTTGTAGCCGCTGATGATC
GphF R675E F	pMAS 425	TCATCAGCGGCTACGAGCCCGAGGATGTCG
GphF R675E R		CGACATCCTCGGGCTCGTAGCCGCTGATGA
GphF H690Y F	pMAS 390	GTGTGATCGTGATCTATGCGCTCCGTGGGGC
GphF H690Y R		GCCCCACGGAGCGCATAGATCACGATCACAC
CurA Q326R F	pMAS 423	GTAAATATCTTGCCAGAGTTGAGAAATCAGGGATTGGGAGATCG

CurA Q326R R		CGATCTCCCAATCCCTGATTTCTCAACTCTGGCAAGATATTTAC
CurA T355S F	pMAS 384	GGTGTCGAAAAAGTTGTGGCAGTATCTCTTTGTCGCAA
CurA T355S R		TTGCGACAAAGAGATACTGCCACAACTTTTTCGACACC
CurA L356R F	pMAS 385	AAAGTTGTGGCAGTAACTCGTTGTCGCAATTATCCAGAC
CurA L356R R		GTCTGGATAATTGCGACAACGAGTTACTGCCACAACTTT
CurA R404K F	pMAS 422	TGAGAAACTGCTTCCTGGTTACAAGCCTAAAGATTGGGAAAACCAG
CurA R404K R		CTGGTTTTCCCAATCTTTAGGCTTGTAACCAGGAAGCAGTTTCTCA
CurA R404E F	pMAS 421	TGAGAAACTGCTTCCTGGTTACGAGCCTAAAGATTGGGAAAACCAG
CurA R404E R		CTGGTTTTCCCAATCTTTAGGCTCGTAACCAGGAAGCAGTTTCTCA
CurA Y419H F	pMAS 405	AGACTTGTGGAGTTCTCGTATCTCACGATATTCAACATCG
CurA Y419H R		CGATGTTGAATATCGTGAGATACGAGAACTCCACAAGTCT

Bold text indicates handles for ligation-independent cloning into expression vectors
All primers are listed 5' to 3'

HisTrap column (GE Healthcare) or 5 mL of Ni-NTA resin (Qiagen) equilibrated with Tris buffer A. After washing with 10 column volumes of Tris buffer A, proteins were eluted with a 5-100% gradient of Tris buffer B (50 mM Tris 7.4, 300 mM NaCl, 10% glycerol, 400 mM imidazole) over 10 column volumes. Single-step purified CurA GNAT was dialyzed into Tris Buffer A for acyl transfer assays. Proteins used for decarboxylation assays and crystallization were further purified by size exclusion chromatography (HiLoad 16/60 Superdex S75) in Tris buffer C (50 mM Tris pH 7.4, 150 mM NaCl, 10% glycerol).

GphF GNAT used for acyl transfer assays was dialyzed into 50 mM BisTris pH 6.5, 20 mM NaCl, 10% glycerol (BisTris buffer A) and loaded on a HiTrap Q anion exchange column (GE Healthcare) equilibrated with BisTris buffer A. GphF GNAT was collected from the Q column flow through and buffer exchanged into Tris buffer C via size exclusion chromatography (HiLoad 16/60 Superdex S75).

A multistep purification was used for CurA GNAT intended for acyl transfer assays. Cell pellets were resuspended in 5 mL Tris buffer D (100 mM Tris pH 7.9, 300 mM NaCl, 10% glycerol, 15 mM imidazole) per 1 g cell pellet with 4 mg DNase, 10 mg lysozyme, and 4 mM

$MgCl_2$ and incubated on ice for 30 min. Cells were lysed by three passes through an Avestin EmulsiFlex-C3 homogenizer and cleared by centrifugation (30,000 x g, 30 min, 4°C). The supernatant was filtered and loaded onto a 5 mL HisTrap column equilibrated with Tris buffer D. Protein was eluted with a gradient of 30-300 mM imidazole. The His-tag was cleaved by overnight incubation with tobacco etch virus (TEV) protease (1:30 ratio of TEV protease: CurA GNAT with 2 mM DTT) and dialyzed into Tris buffer D. The digested protein was passed over a second 5 mL HisTrap column. Tag-free CurA GNAT was collected from the flow through and further purified by size exclusion chromatography (HiLoad 16/60 Superdex S200) in Tris buffer D.

GphF MT1-GNAT used for acyl transfer assays was purified as described previously[119] and then diluted into Hepes buffer A (50 mM Hepes pH 7.4, 15 mM NaCl, 10% glycerol). GphF MT1-GNAT was then loaded onto a HiTrap Q Column equilibrated with Hepes buffer A, washed with 10 column volumes of Hepes buffer A and eluted with a 15-1 M linear gradient of NaCl over 20 column volumes. GphF MT-GNAT was buffer exchanged into Hepes buffer B (50 mM Hepes pH 7.4, 150 mM NaCl, 10% glycerol) via size exclusion chromatography (HiLoad 16/60 Superdex S200).

AprA apo-ACP was produced and purified as described previously[119]. A trace metals mix was included in the growth medium to inhibit addition of the phosphopantetheinyl post translational modification by endogenous *E. coli* enzymes[168]. CurA ACP was produced and purified identically to AprA ACP. *M. bouillonii* FabD[120] and AprA MT1-ΨGNAT[119] were produced and purified as described previously.

Production of acyl-ACPs

AprA and CurA holo-, malonyl- (Mal-), and methylmalonyl- (MeMal-)ACPs were produced by incubating 180 μM apo-ACP with 20 μM *Streptomyces verticillus* phosphopantetheinyl transferase (SVP)[142], and ~0.65 mM CoA, Mal-CoA, or MeMal-CoA in 50 mM Tris pH 7.4, 150 mM NaCl, 10% glycerol, 20 mM $MgCl_2$ for 4 hr at 30°C. The ACP was purified from the reaction mixture by size exclusion chromatography (HiLoad 16/60 Superdex S75) equilibrated with Tris buffer E (100 mM Tris pH 7.4, 250 mM NaCl, 5% glycerol, 5 mM tris(2-carboxyethyl)phosphine (TCEP).

As dimethylmalonyl (Me_2Mal)-CoA is not commercially available, AprA MT1[119] was used to convert the AprA and CurA MeMal-ACPs to the Me_2Mal-ACPs. Reaction mixtures for AprA ACP (~3 mL total in 130 μL aliqots) containing 300 μM AprA MeMal-ACP, 150 μM AprA MT1-

ΨGNAT and 6 mM *S*-adenosylmethionine (SAM) in 50 mM Hepes pH 7.4, 150 mM NaCl, 3 mM $Fe(NH_4)_2(SO_4)_2$ were incubated 5.5 hr at 30°C. Reaction mixtures for CurA ACP (~4 mL total in 50 µL aliqots) containing 150 µM CurA MeMal-ACP, 150 µM AprA MT1-ΨGNAT and 6 mM SAM in 50 mM Hepes pH 7.4, 150 mM NaCl, 3 mM $Fe(NH_4)_2(SO_4)_2$ were incubated 6 hr at 30°C. Me_2Mal-ACPs were purified from the reaction mixtures by size exclusion chromatography (HiLoad 16/60 Superdex S75) equilibrated with Tris buffer E. Intact protein mass spectra coupled with the Ppant ejection assay[150, 169] was used to validate the conversion of MeMal-ACP to Me_2Mal-ACP.

Decarboxylation enzyme assays

As GphF ACP_L could not be produced in soluble form, AprA ACP_L from the apratoxin A biosynthetic pathway[25] was used as a surrogate ACP. Reaction mixtures (80-100 µL) containing 100 µM AprA Mal-, MeMal-, or Me_2Mal-ACP and 10 µM GphF GNAT in 50 mM HEPES pH 7.4, 150 mM NaCl were incubated at 30°C. 10 µL of the reaction mixture was collected at various time points and quenched with 1% (v/v) formic acid. 0.25 µL of reaction mixtures were subjected to LC/MS analysis.

Reaction mixtures (10 µL) containing 100 µM AprA acyl-ACP and 10 µM GphF GNAT variants in 50 mM HEPES pH 7.4, 150 mM NaCl were incubated at 30°C for 5 min (Me_2Mal-ACP), 8 hr (MeMal-ACP), or 24 hr (Mal-ACP). Reactions were quenched with 1% (v/v) formic acid. 0.25 µL of reaction mixtures were subjected to LC/MS analysis.

Reaction mixtures (80-100 µL) containing 100 µM CurA Mal-, MeMal-, or Me_2Mal-ACP and 10 µM CurA GNAT in 50 mM HEPES pH 7.4, 150 mM NaCl were incubated at 30°C. 10 µL of the reaction mixture was collected at various time points and quenched with 1% (v/v) formic acid. 0.1 µL of reaction mixtures were subjected to LC/MS analysis.

Reaction mixtures (10 µL) containing 100 µM CurA acyl-ACP and 10 µM CurA GNAT variants in 50 mM HEPES pH 7.4, 150 mM NaCl were incubated 6 hr (Me_2Mal-ACP), 7.5 min (MeMal-ACP), or 3 min (Mal-ACP). Reactions were quenched with 1% (v/v) formic acid. 0.1 µL of reaction mixtures were subjected to LC/MS analysis.

Acyl transfer enzyme assays

Acyl transfer reactions (70 µL) containing 100 µM AprA or CurA holo-ACP, 10 µM CurA GNAT or GphF GNAT, 0.85 mM of Mal-, MeMal-, acetyl-, propionyl-, or isobutyryl-CoA in 50

mM Hepes pH 7.4, 150 mM NaCl were incubated at 30°C. At various time points 10 µL aliquots were removed and quenched with 1% (v/v) formic acid. 0.25 µL of GphF GNAT reactions and 0.1 µL of CurA GNAT reactions were used for LC/MS analysis.

GphF MT_L-GNAT acyl transfer reactions (70 µL) containing 100 µM AprA holo-ACP, 10 µM GphF MT_L-GNAT, 0.85 mM of Mal-CoA, 0.54 mM SAM, 50 mM Hepes pH 7.4, 150 mM NaCl, 0.5 mM $FeH_8N_2O_8S_2$ were incubated at 30°C. At various time points 10 µL aliquots were removed and quenched with 1% (v/v) formic acid. 0.25 µL of the reaction mixtures were subjected to LC/MS analysis.

M. bouillonii FabD reaction mixtures (120 µL) containing 100 µM CurA holo-ACP, 25 nM FabD, and 0.85 mM Mal-CoA in 50 mM Hepes 7.4, 150 mM NaCl were incubated at 30°C. At various time points 10 µL aliquots were removed and quenched with 1% (v/v) formic acid. 0.1 µL of the reaction mixtures were subjected to LC/MS analysis.

LC-MS analysis

An Agilent Q-TOF 6545 was used to analyze reaction mixtures using the phosphopantetheine (Ppant) ejection method[150, 169]. Samples underwent reverse phase HPLC (Phenomenex Aeris widepore C4 column 3.6 µM, 50 x 2.10 mm) in H_2O with 0.2% (v/v) formic acid at a flow rate of 0.5 mL min^{-1}. Protein was eluted over a 4 min linear gradient of 5-100% acetonitrile with 0.2% (v/v) formic acid. Samples were subjected to the following conditions: fragmentor voltage, 225 V; skimmer voltage, 25 V; nozzle voltage, 1000 V; sheath gas temperature, 350 °C; drying gas temperature, 325 °C. Data was processed using MassHunter Qualitative Analysis Software (Agilent). The relative abundances of Ppant ejection fragments for substrates and products were used to calculate the percent of total ACP species. In acyl transfer reactions with active GNAT domains, abundances of intact and decarboxylated acyl-ACP species were combined to calculate percent acyl group transfer.

Protein crystallization and structure determination

GphF GNAT (residues 498-705, with additional N-terminal His-tag) was crystallized at 20°C by sitting drop vapor diffusion from a 2:1 µL mixture of protein stock (20.7 mg mL^{-1} GphF GNAT in Tris buffer C) and reservoir solution (30 - 35% PEG 3350, 0.23 - 0.30 M ammonium acetate, 0.10 M Bis-tris HCl pH 5.5). Crystals of thin-blade morphology grew overnight, often in clusters, and could be divided into apparently single crystals for data collection. Crystals were

cryoprotected with well solution supplemented with 15% glycerol and flash cooled in liquid N_2. For the isobutyryl-CoA complex, crystals were soaked with 2.5 mM isobutyryl-CoA for 24 hr prior to harvesting.

Diffraction data were collected at 100 K on GM/CA beamline 23ID-B at the Advanced Photon Source (APS) at Argonne National Laboratory (Argonne, IL). Data were processed using XDS[128]. The thin-blade morphology of the crystals limited the data quality. Diffraction patterns were streaky and moderately anisotropic (diffraction limit for GphF GNAT: d_{min} = 3.3 Å along a^*, and 2.6 Å along b^* and c^*; isobutyryl-CoA complex: d_{min} = 5.0 Å along a^* and 2.8 Å along b^* and c^*). Probing the crystal with a 10-µm beam in raster cells did not identify local regions with better spot shapes or lower anisotropy. The structure of GphF GNAT was solved by molecular replacement using Phaser[172] through the Phenix Software Suite[130] with CurA GNAT (2REE)[31] as the starting model. Crystals of the isobutyryl complex were isomorphous with crystals of the free enzyme. Iterative rounds of model building and refinement were carried out using Coot[132] and Phenix.refine[173] with automated translation/liberation/screw group selection. Active site cavities were computed using CAVER[135]. Structures were validated with MolProbity[133], and structure figures were prepared with PyMol[136]. Sequence alignments were prepared using Clustal[137] through Jalview[138].

Table 6.2 GphF GNAT crystallographic information

Protein	GphF GNAT	GphF GNAT
Ligand		Isobutyryl-CoA
Data Collection		
Space group	*C*2 2 2_1	*C*2 2 2_1
Unit cell, a,b,c (Å)	138.3, 145.7, 78.0	134.4, 145.5, 77.2
X-ray source	APS 23ID-B	APS 23ID-B
Wavelength (Å)	1.033	1.033
d_{min} (Å)	2.59 (2.68-2.59) [a]	2.79 (2.89-2.79)
R-merge	0.1847 (3.16)	0.1762 (1.97)
Inner-shell R-merge	0.047 (7.67 Å) [b]	0.035 (8.28 Å)
Avg I/σ(I)	11.7 (0.9)	9.87 (1.10)
Completeness (%)	99.6 (97.3)	99.3 (95.7)
Multiplicity	13.1 (13.4)	6.8 (6.9)
Total observations	327,190 (31,990)	128,891 (12,396)
Wilson B factor ($Å^2$)	76.9	69.2
$CC_{1/2}$	0.998 (0.463)	0.997 (0.644)
CC*	1.00 (0.795)	0.999 (0.885)
Refinement		
Data range (Å)	42.19-2.59	42.8-2.79
Reflections	24,860	19,043
R_{work}/R_{free} (%)	23.0/26.8	23.2/27.5
Non-hydrogen atoms (#)	3,011	3,070
protein	2,934	2,934
ligands	35	122
water	42	14
Amino acid residues	390	392
Deviation from ideality		
bond lengths (Å)	0.004	0.002
bond angles (°)	0.98	0.42
Average B-factor ($Å^2$)	91.6	103.5
protein	91.5	103.4
ligands	112.8	111.6
solvent	76.4	66.7
Ramachandran plot		
favored (%)	98.2	96.9
allowed (%)	1.8	3.1
outliers (%)	0	0

[a]values in parentheses designate outer shell
[b]d_{min} inner shell

Table 6.3 Scaling statistics for GphF GNAT

Resolution Limit (Å)	Number of Reflections						
	Observed	Unique	Possible	Completeness	R-Factor	I/σ	$CC_{1/2}$
7.67	12623	1046	1054	99.20%	4.70%	46.62	99.9
5.46	21662	1754	1755	99.90%	8.90%	29.62	99.8
4.47	28720	2220	2221	100.00%	10.10%	27.42	99.8
3.88	35965	2593	2594	100.00%	15.20%	20.59	99.7
3.47	40469	2925	2925	100.00%	24.30%	13.18	99.5
3.17	40828	3233	3233	100.00%	48.50%	6.34	98.5
2.93	43609	3485	3485	100.00%	121.70%	2.44	91.7
2.75	50444	3743	3743	100.00%	195.50%	1.5	74.6
2.59	52918	3893	3962	98.30%	299.80%	0.96	49.8
Total	327238	24892	24972	99.70%	18.40%	11.73	99.8

Table 6.4 Scaling statistics for GphF GNAT in complex with isobutyryl-CoA

Resolution Limit (Å)	Number of Reflections						
	Observed	Unique	Possible	Completeness	R-Factor	I/σ	$CC_{1/2}$
8.28	4965	804	815	98.70%	3.50%	40.61	99.9
5.89	8479	1346	1347	99.90%	7.90%	21.14	99.8
4.82	11521	1697	1697	100.00%	11.20%	17.05	99.7
4.18	14166	2007	2008	100.00%	11.30%	17.14	99.7
3.74	14845	2251	2251	100.00%	18.40%	11.35	99.3
3.42	15978	2474	2474	100.00%	28.90%	7.39	98.7
3.17	18243	2675	2675	100.00%	56.90%	3.88	95.4
2.96	20097	2864	2866	99.90%	109.10%	2.11	88.1
2.79	20612	2962	3038	97.50%	169.50%	1.3	73.7
Total	128906	19080	19171	99.50%	17.60%	9.88	99.8

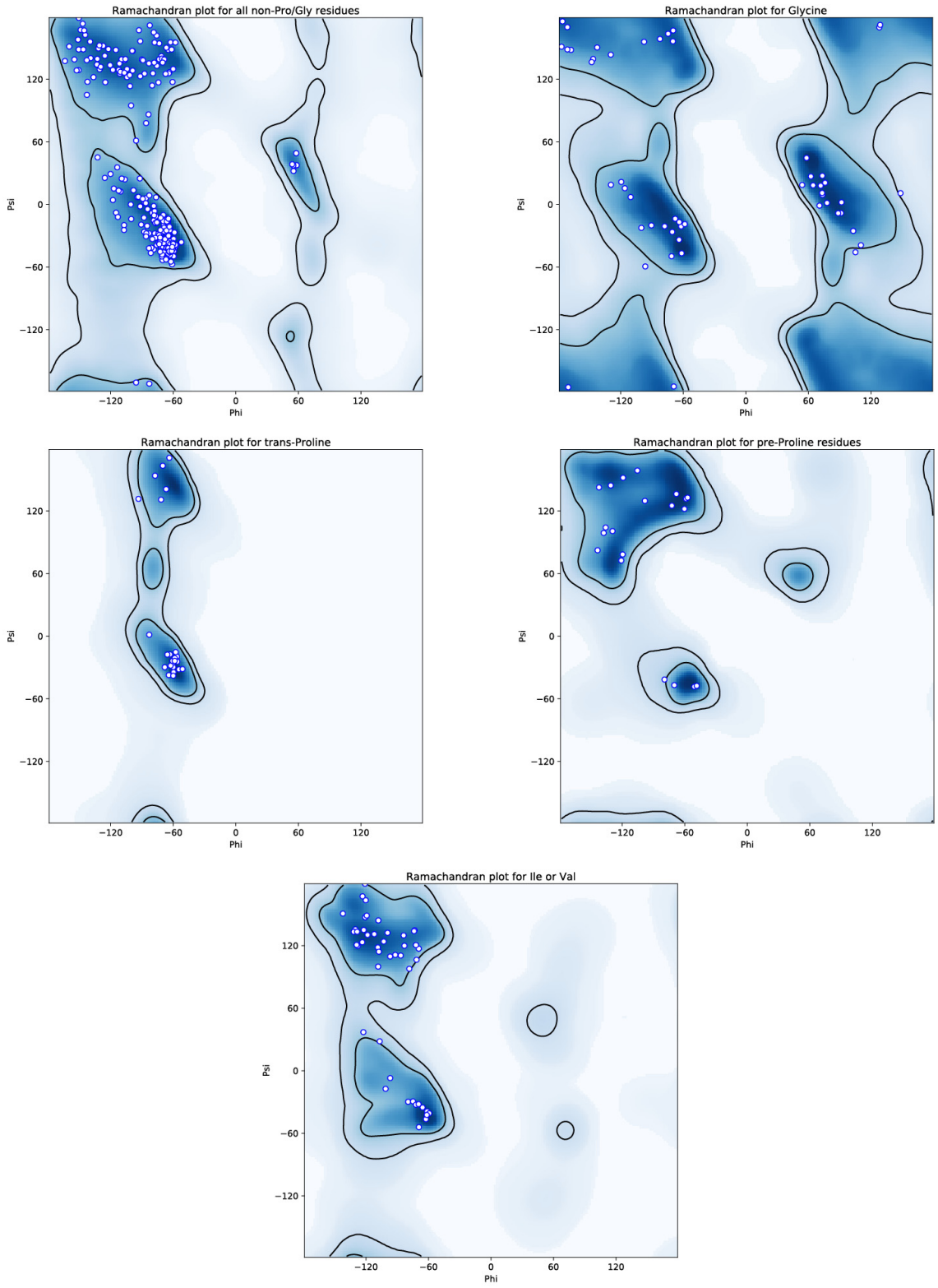


Figure 6.3 Ramachandran plots for GphF GNAT
Ramachandran plots of final refined model of GphF GNAT. Plots were generated using MolProbity[133].

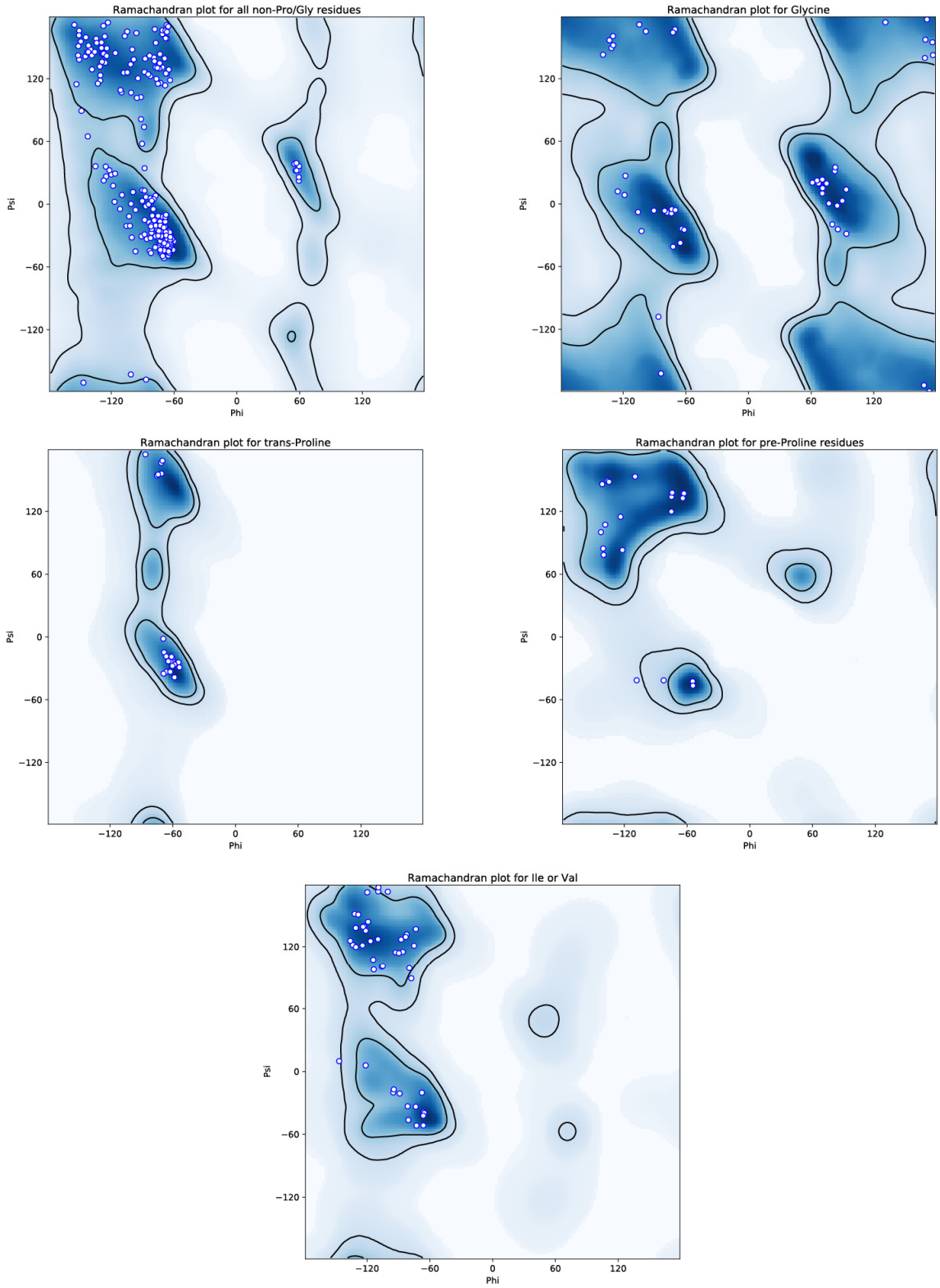


Figure 6.4 Ramachandran plots for GphF GNAT in complex with isobutyryl-CoA
Ramachandran plots of final refined model of GphF GNAT with bound isobutyryl-CoA. Plots were generated using MolProbity[133].

Results and Discussion

GNAT-like enzymes encode strict substrate specificity

The metal-dependent MT_L in PKS initiation modules can produce both monomethylated (MeMal-ACP)[219] and dimethylated (Me_2Mal-ACP)[119] products. We hypothesized that the GNAT-like domain may act as a gatekeeper, acting selectively after MT_L has installed the pathway-specific number of methyl groups. The structure and pathway annotation of gephyronic acid[21] as well as initial biochemical characterization[119] indicate that the GphF MT_L produces Me_2Mal-ACP via a MeMal-ACP intermediate (Figure 6.1d). Thus, the GphF GNAT displays selectivity for Me_2Mal-ACP over MeMal-ACP to ensure production of the isobutyryl starter unit.

As GphF ACP_L was insoluble, AprA ACP_L, the ACP from the initiation module of the apratoxin A pathway[25], was used as a surrogate ACP to assay GphF GNAT function. AprA ACP was loaded with Mal- and MeMal-CoA substrates via the promiscuous phosphopantetheinyl transferase SVP[142] to produce Mal-ACP, the substrate of the GphF MT_L, and its MeMal-ACP intermediate (Figure 6.1d). MeMal-ACP was converted to Me_2Mal-ACP via AprA MT_L, the apratoxin A initiation module methyltransferase[119], as Me_2Mal-CoA is commercially unavailable and challenging to access in sufficient quantities synthetically. Relative rates of decarboxylation of the ACP-linked substrates were measured using the mass-spectrometry-based Ppant ejection assay[150, 169]. GphF GNAT displayed a strong selectivity for decarboxylation of Me_2Mal-ACP relative to MeMal-ACP (100-fold) or Mal-ACP (barely detectable within a 48-hours) (Figure 6.5a, Figure 6.6). Thus, GphF GNAT requires the action of MT_L prior to decarboxylation and is selective for the Me_2Mal-ACP product over the MeMal-ACP intermediate to prime the gephyronic acid assembly line with the isobutyryl-ACP starter group.

We were curious to examine whether other PKS GNAT-like domains exhibit a similar degree of substrate selectivity. The CurA GNAT[31] is of particular interest as the CurA initiation module includes the N-terminal lid region (formerly known as an adaptor region) of a MT_L, but has no SAM binding core (Figure 6.1a), suggesting that CurA evolved from a module containing an MT_L[119]. CurA GNAT displayed modest selectivity with a ~5-fold preference for its natural substrate Mal-ACP over MeMal-ACP (Figure 6.5b), in agreement with previously reported values[31]. Interestingly, CurA GNAT also accommodated the Me_2Mal-ACP substrate, but produced isobutyryl-ACP at a ~500-fold slower rate than acetyl-ACP (Figure 6.5b). CurA GNAT

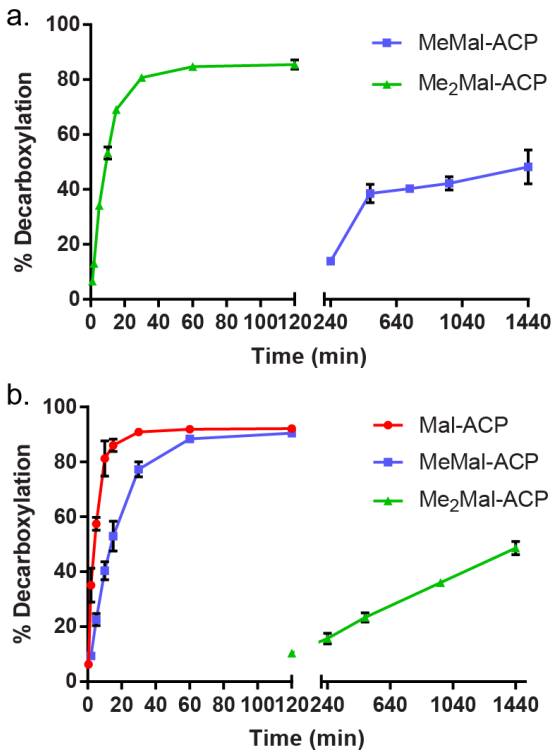


Figure 6.5 GphF and CurA GNAT decarboxylation activity
Decarboxylation of various acyl-ACP substrates by **a.** GphF GNAT and **b.** CurA GNAT. Activity was monitored via the Ppant ejection assay[150, 169]. Error bars represent triplicate experiments and, in some cases, are too small to be visible. GphF GNAT did not decarboxylate Mal-ACP in a 48 hour reaction time.

decarboxylation of MeMal-ACP is not likely to result in mispriming of curacin biosynthesis with a propionyl-ACP starter unit as many bacteria, including cyanobacteria, lack propionyl-CoA carboxylase and do not produce MeMal-CoA at levels sufficient for secondary metabolite biosynthesis[246].

GphF GNAT structure

In order to understand the substrate selectivity of PKS GNAT-like domains for decarboxylation, we solved a 2.6 Å crystal structure of the GphF GNAT and a 2.8 Å structure in complex with an isobutyryl-CoA product mimic (Figure 6.7, Table 6.2). Like the CurA GNAT, GphF GNAT displays the conserved GNAT fold comprised of a β-sheet core flanked by α-helices. The GNAT fold is marked by a "β-bulge" (GphF residues Leu588, Gly589) in β4, creating the V-

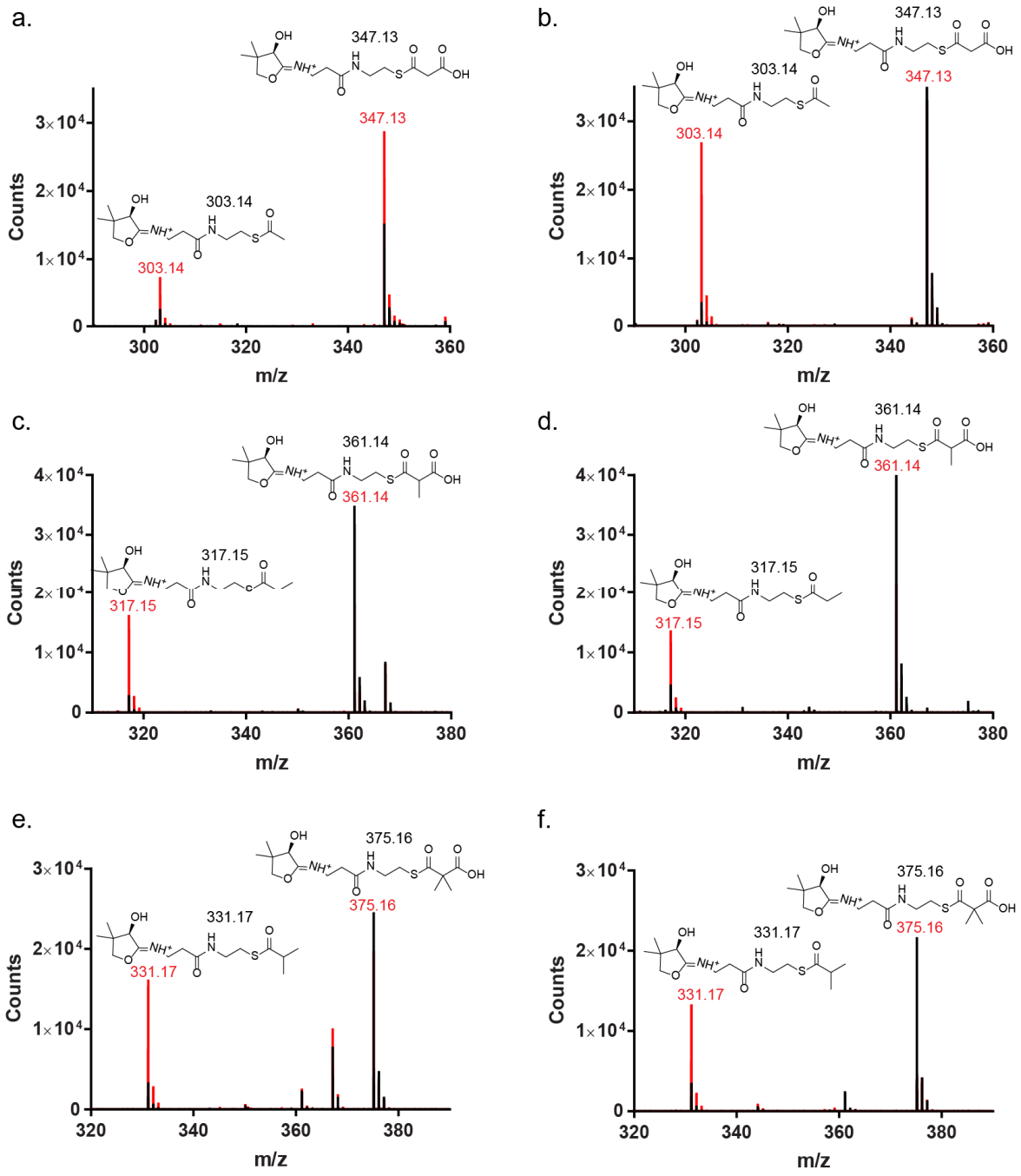


Figure 6.6 Representative mass spectra of decarboxylation reactions
Decarboxylation of Mal-ACP:**a.** GphF GNAT, 48 hr reaction time, **b.** CurA GNAT, 3 min reaction time; MeMal-ACP: **c.** GphF GNAT, 8 hr reaction time, **d.** CurA GNAT, 7.5 min reaction time; Me_2Mal-ACP: **e.** GphF GNAT, 5 min reaction time, **f.** CurA GNAT, 8 hr reaction time. Spectra of no enzyme controls are shown in black; reactions are shown in red. Calculated m/z values are listed in black above structures of Ppant ejection fragments. Observed m/z values are indicated in red. AprA ACP, used as a surrogate ACP for GphF GNAT reactions, had a contaminating species at 367 Da. The same ratio of malonyl- and acetyl-Ppant ejection fragments to holo-Ppant ejection fragments are observed in the no enzyme control and reaction in a.

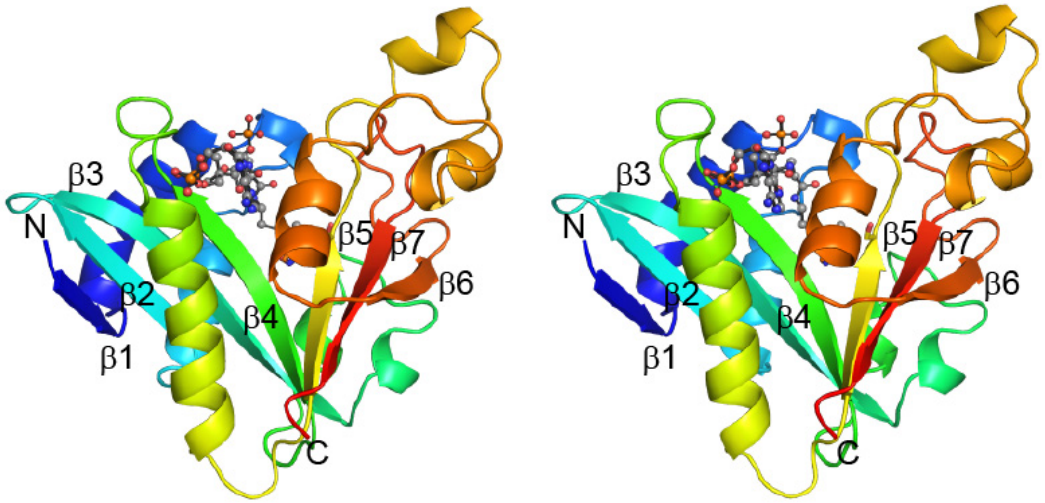


Figure 6.7 GphF GNAT structure
The protein is colored as a rainbow from blue (N-terminus) to red (C-terminus), shown in stereo. Isobutyryl-CoA is (ball-and-stick form with atomic colors: C, gray; O, red; N, blue; S, yellow) binds to a V-shaped cleft between β4 (green) and β5 (yellow).

shaped Ppant binding cleft between the parallel β4 and β5 strands (Figure 6.7)[247]. PKS GNAT-like domains are characterized by an insertion after β5 and by an extended C-terminus that inserts a strand between β5 and β6 (Figure 6.7). The isobutyryl-CoA diphosphate is stabilized by backbone amides of the phosphate binding loop between β4 and α5 (GphF residues 597-602), as in other GNAT family members (Figure 6.7, Figure 6.2).

GNAT active site and catalysis

At the substrate entrance site, the isobutyryl-CoA Ppant threads along β4, mimicking an additional β-strand (Figure 6.7). The thioester is positioned at the base of the V-shaped cleft between β4 and β5 near conserved Thr/Ser and His residues (GphF Ser626, His660; CurA Thr355, His 389), which are essential for decarboxylation by CurA GNAT (Figure 6.8)[31]. The binding pocket for acceptor substrates of GNAT acyltransferases lies beyond the isobutyryl-Ppant. In PKS GNAT-like domains, conserved Trp and Arg side chains (GphF Trp520, Arg 675; CurA Trp 249, Arg404) are at the proximal end of the acyl pocket.

The role of GphF active site residues in decarboxylation of Me_2Mal-ACP and MeMal-ACP was probed via site-directed mutagenesis (Figure 6.8). In addition, we tested all GphF GNAT variants for the ability to decarboxylate Mal-ACP. Like the CurA GNAT, the GphF GNAT conserved active site His (GphF H660) and Ser/Thr (GphF Ser626) are essential for

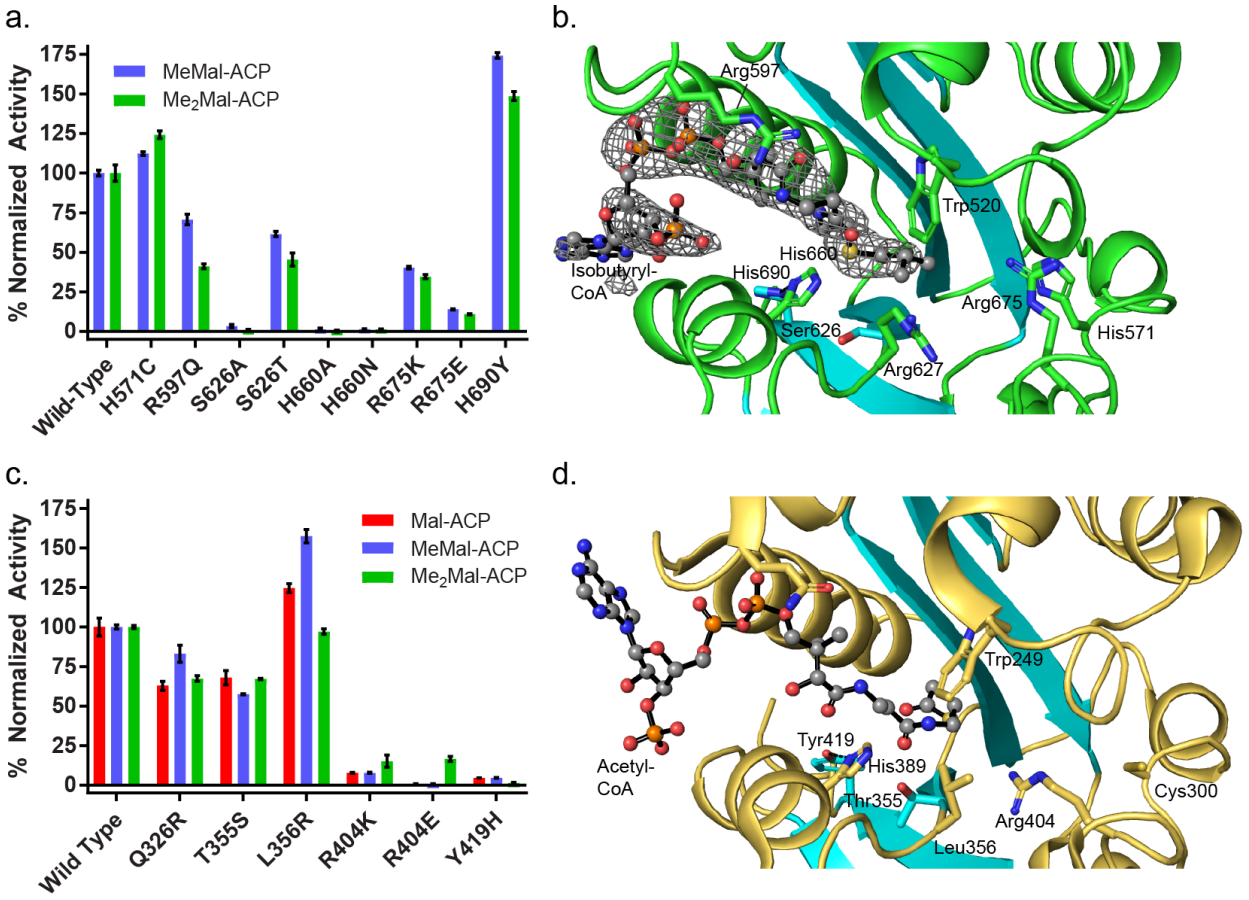


Figure 6.8 Probing decarboxylation via site directed mutagenesis.
a. GphF GNAT decarboxylation reactions. **b.** GphF GNAT active site with bound isobutyryl-CoA. *Fo*-*Fc* omit density for isobutyryl-CoA is shown in gray mesh at 3σ contour. Ser626 and His660 are the catalytic amino acids. **c.** CurA GNAT decarboxylation reactions. **d.** CurA GNAT active site with bound acetyl-CoA[31]. Thr355 and His389 are the catalytic amino acids. Error bars represent triplicate experiments and, in some cases, are too small to be visible. Bound CoAs are shown in ball-and-stick form. Amino acids subjected to mutagenesis are shown in sticks with atomic coloring.

decarboxylation of both Me_2Mal-ACP and MeMal-ACP (Figure 6.2, Figure 6.8a, b). Like the wild type GNAT, no decarboxylation of Mal-ACP was detected in a 24 hour incubation. Next, the conserved active site Arg (GphF R675) was substituted with Lys and Glu, which reduced product formation by fourfold (R675K) and tenfold (R675E) (Figure 6.8 a, b). Analogous substitutions in CurA Arg 404 were similarly deleterious, validating the role of the conserved Arg in decarboxylation (Figures 4c and 4d). Given the near elimination of product formation for the charge-reversal substitutions (GphF R675E and CurA R404E) and the substantial decrease for the more conservative substitutions (GphF R675K and CurA R404K), we hypothesize that the positively charged Arg interacts with the Me_2Mal- or Mal-ACP carboxylate. The Arg675

guanidinium is well positioned to interact with the carboxylate of a dimethylmalonyl substrate in a model based on the structure of the GphF GNAT complex with isobutyryl-CoA (Figure 6.9). The essential active site His and Thr/Ser are the best candidates to stabilize an enolate intermediate formed during decarboxylation and to re-protonate the carbanion upon collapse of the enolate. In some GNAT superfamily members that catalyze acyl transfer, two backbone amides at the bulge in β4 point into the active site and form an oxyanion hole to stabilize the enolate intermediate. At the bulge of the GphF and CurA GNATs, backbone carbonyl groups and not amides point into the active site.

We next examined amino acids in the Ppant binding cleft and acyl group pocket that differ in GphF and CurA GNAT and evaluated their effects on catalysis and substrate selectivity. The essential side chains Thr and His in the decarboxylation site of PKS GNAT-like enzymes (CurA Thr355, His389) are more common than the Ser/His pair (GphF Ser626, His660) (Figure 6.2, Figure 6.8a, b). For GNAT sequences where the pathway product is known, only the GphF initiation module produces an isobuturyl starter group. We hypothesized that the smaller active site Ser may provide more space for a dimethylmalonyl than would the bulkier Thr. However, the GphF S626T and CurA T355S substitutions did not appear to influence selectivity for Me_2Mal-ACP and had only modest effects on decarboxylation of all substrates (Figure 6.8 a, b).

In CurA GNAT, the essential His (CurA His389) is positioned by a hydrogen bond to the hydroxyl of a Tyr (CurA Tyr419), which is conserved in all PKS GNAT-like domains except the GphF GNAT where His690 is just beyond hydrogen bonding distance to the catalytic His660 (Figure 6.2, Figure 6.8b, d). A CurA Y418H substitution eliminated decarboxylation activity, whereas the reciprocal GphF GNAT substitution (GphF H690Y) enhanced decarboxylation of Me_2Mal-ACP and MeMal-ACP (Figure 6.8a, c) but did not alter substrate selectivity. Tyr is more effective than His at polarizing the catalytic His660 imidazole, which may account for the enhanced reactivity. Additional features of the GphF GNAT active site environment must modulate the reactivity of the catalytic His (His660) to compensate for the lack of a Tyr.

GphF Arg597 is found at the entrance of the Ppant binding cleft (Figure 6.8b) whereas the analogous position is glutamine in CurA (CurA Q326) and other PKS GNAT-like domains (Figure 6.2, Figure 6.8d). The GphF Arg597 guanidinium and CurA Gln326 amide both stack with a Ppant amide group. The GphF R597Q variant had variable effects on decarboxylation, decreasing the production of propionyl-ACP by 30% and isobuturyl-ACP by 60% (Figure 6.8a). The analogous

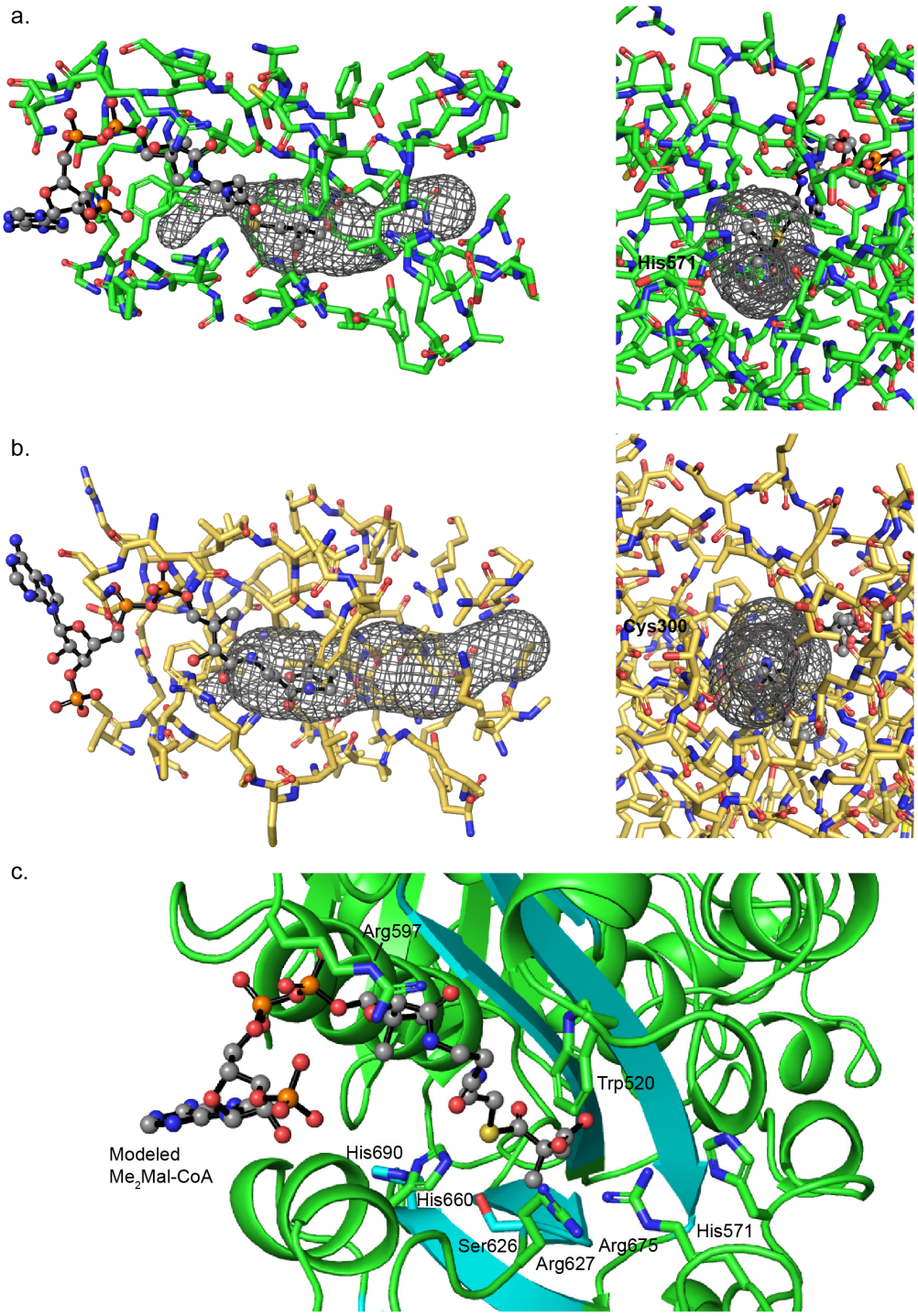


Figure 6.9 Active site pockets of GNAT-like enzymes and modeled Me_2Mal-CoA complex
a. GphF, **b.** CurA. Gray mesh depicts tunnel calculated using CAVER[135]. Images on the left and right differ by 90°. The GphF active site pocket is wider near the Ppant thiol to accommodate the Me_2Mal-ACP substrate. Access to solvent through the distal end of the acyl binding pocket (right images) is partially occluded in the GphF domain but open to solvent in CurA. **c.** Me_2Mal modeled into GphF GNAT based on the position of isobutyryl-CoA. The terminal carboxylate can be placed near conserved Arg675

CurA Q326R variant decreased decarboxylation of Mal-ACP, MeMal-ACP, and Me_2Mal-ACP less than twofold (Figure 6.8c). Thus, neither Arg nor Gln appear to play a role in substrate selectivity or catalysis.

A major difference between the GphF and CurA GNAT active sites is the presence of an Arg in GphF (Arg627) at the position of a Leu in CurA (Leu356) (Figure 6.8b, d). Leu356 is near the conserved active site Trp (CurA W249) and contributes to narrowing the acyl binding pocket relative to the GphF GNAT. Substitution of Arg for Leu356 in CurA enhanced decarboxylation of Mal-ACP and MeMal-ACP substrates, but did not affect the slow decarboxylation of Me_2Mal-ACP, and thus does not contribute to substrate selectivity (Figure 6.8c). The corresponding GphF R627L variant (Figure 6.8b) was unstable.

The acyl binding pocket of GNAT acyltransferases extends to the surface of the protein, opening to solvent between α2 and the β3-β4 loop. A PEG molecule from the crystallization solution occupies the pocket in the GphF GNAT free enzyme structure. The corresponding pocket was proposed to accommodate the acyl-accepting ACP Ppant in CurA GNAT[31], however access through the distal end of the pocket is limited in GphF due to the bulkier His571 compared to Cys300 in CurA GNAT (Figure 6.8b, d, Figure 6.9). Amino acids in this region are among the least conserved in PKS GNAT-like domains. The variability does not suggest a common Ppant pocket for ACP acceptors, as would be expected in acyl transfer enzymes. In some PKS GNAT-like enzymes, access to solvent may be further occluded by a larger aromatic side chains at this position (Figure 6.2). A GphF H571C variant had little effect on decarboxylation (Figure 6.8a), indicating that access to the active site through the acyl binding pocket is not crucial for decarboxylation of the ACP-linked substrates.

GphF GNAT does not catalyze acyl transfer

The proposed reaction scheme for the GphF initiation module requires the transfer of malonate from CoA to ACP prior to the methylation reactions by MT_L. This ensures that the resulting Me_2Mal product is non-diffusable and available for further processing by subsequent enzymes in the gephyronic acid pathway. GphF GNAT is a candidate to provide the malonyl transfer step as it was unable to decarboxylate malonyl-ACP prior to methylation by GphF MT_L (Figure 6.1d). However, we detected malonyl transfer activity by GphF GNAT at very low levels, which varied among preparations of the enzyme, unlike the consistently high levels of decarboxylation activity.

The low levels and prep-to-prep variation suggested that the observed acyl transfer activity may be due to a contaminating enzyme from the highly active fatty acid biosynthetic machinery of the *E. coli* heterologous expression host, which has been reported to support transfer of malonyl from CoA to PKS ACPs[196]. To separate GphF GNAT (pI 7.3) from the potential contaminants FabD (pI 5.0) and FabH (pI 5.1), the *E. coli* fatty acid synthase enzymes that could support acyl transfer, we added an additional ion exchange chromatography step to the purification protocol. The stringently purified GphF GNAT did not support acyl transfer to ACP from a panel of acyl-CoA donors, including Mal-CoA and isobutyryl-CoA (Figure 6.10a, Figure 6.11), but it retained a high rate of Me_2Mal-ACP decarboxylation (Figure 6.12a). Thus, we conclude that the GphF GNAT-like enzyme is not an acyltransferase. Acyl transfer reactions with GphF MT_L-GNAT and Mal-CoA showed very slow loading of malonyl onto holo-ACP (Figure 6.10a). The suspiciously slow acyl transfer rate for GphF MT_L could also be a result of lingering contaminants from heterologous expression, as the GphF MT_L-GNAT has a pI of 6.3 and is less amenable to stringent purification by ion exchange chromatography.

The lack of acyl transfer activity for GphF GNAT prompted us to reinvestigate CurA GNAT acyl transfer, which was ~780 fold slower than decarboxylation[31]. CurA GNAT underwent additional affinity chromatography and size exclusion purification steps, as CurA GNAT is not amenable to ion exchange (pI 5.5). The multistep-purified CurA GNAT diminished acyl transfer activity compared to the single-step purified enzyme used previously (2% *vs*. 10% for acetyl-CoA, Figure 6.10b, Figure 6.13), but it retained rapid decarboxylation activity (Figure 6.12). The native malonyl acyl transferase (FabD) of fatty acid biosynthesis in the producing organism is an attractive candidate to provide the initial malonyl acyl transfer step for PKS initiation modules containing a GNAT-like decarboxylation domain, as shown in the apratoxin A biosynthetic pathway[120]. We tested this with the *Moorea bouillonii* (*M. bouillonii)* FabD, which is 97% identical to FabD from the curacin A producer *M. producens*. *M. bouillonii* FabD rapidly transferred malonyl from CoA to CurA ACP (65% transfer in 20 min with 25 nM FabD compared to 8% in 4 hr with 10 μM GNAT, Figure 6.10c, Figure 6.13).

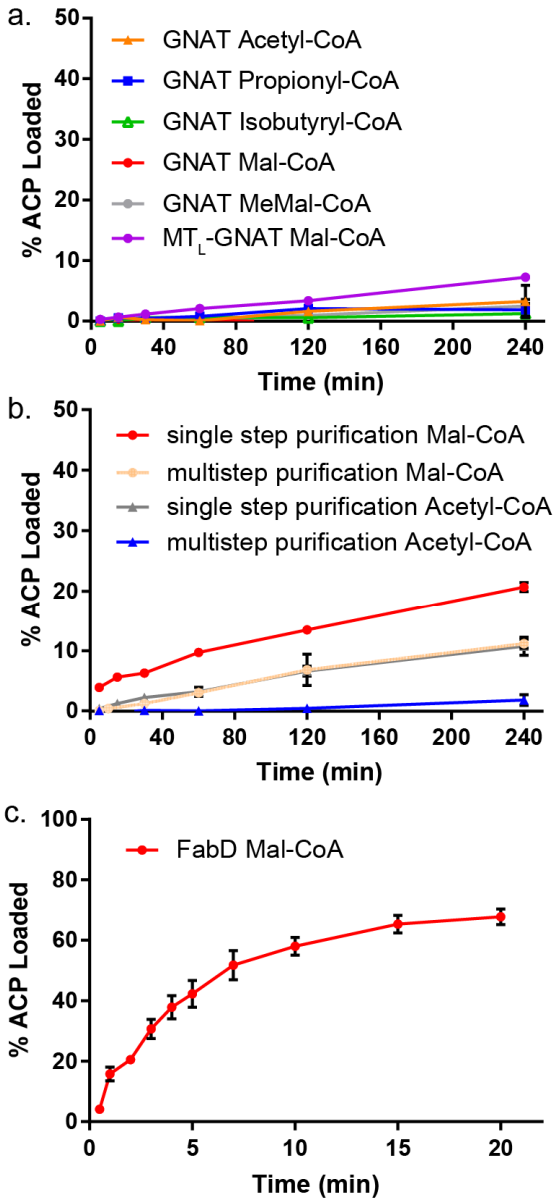


Figure 6.10 GphF and CurA GNAT acyl transfer assays
a. GphF GNAT and GphF MT$_L$-GNAT reactions with acyl-CoAs and AprA holo-ACP. Virtually no acyl group transfer to the ACP was detected with GphF GNAT. GphF MT$_L$-GNAT supported low levels of malonyl transfer. **b.** CurA GNAT reactions with Mal-CoA or acetyl-CoA and CurA holo-ACP. The apparent acyl transfer activity of CurA GNAT preparations diminished with additional purification stops whereas decarboxylation activity was unaffected. **C.** *M. bouillonii* FabD reactions with Mal-CoA and CurA holo-ACP. *M. bouillonii* FabD rapidly transferred malonyl from CoA to CurA ACP at a 1:4000 molar ratio of FabD to ACP. Activity was monitored via the Ppant ejection assay[150, 169]. Error bars represent triplicate experiments and, in some cases, are too small to be visible.

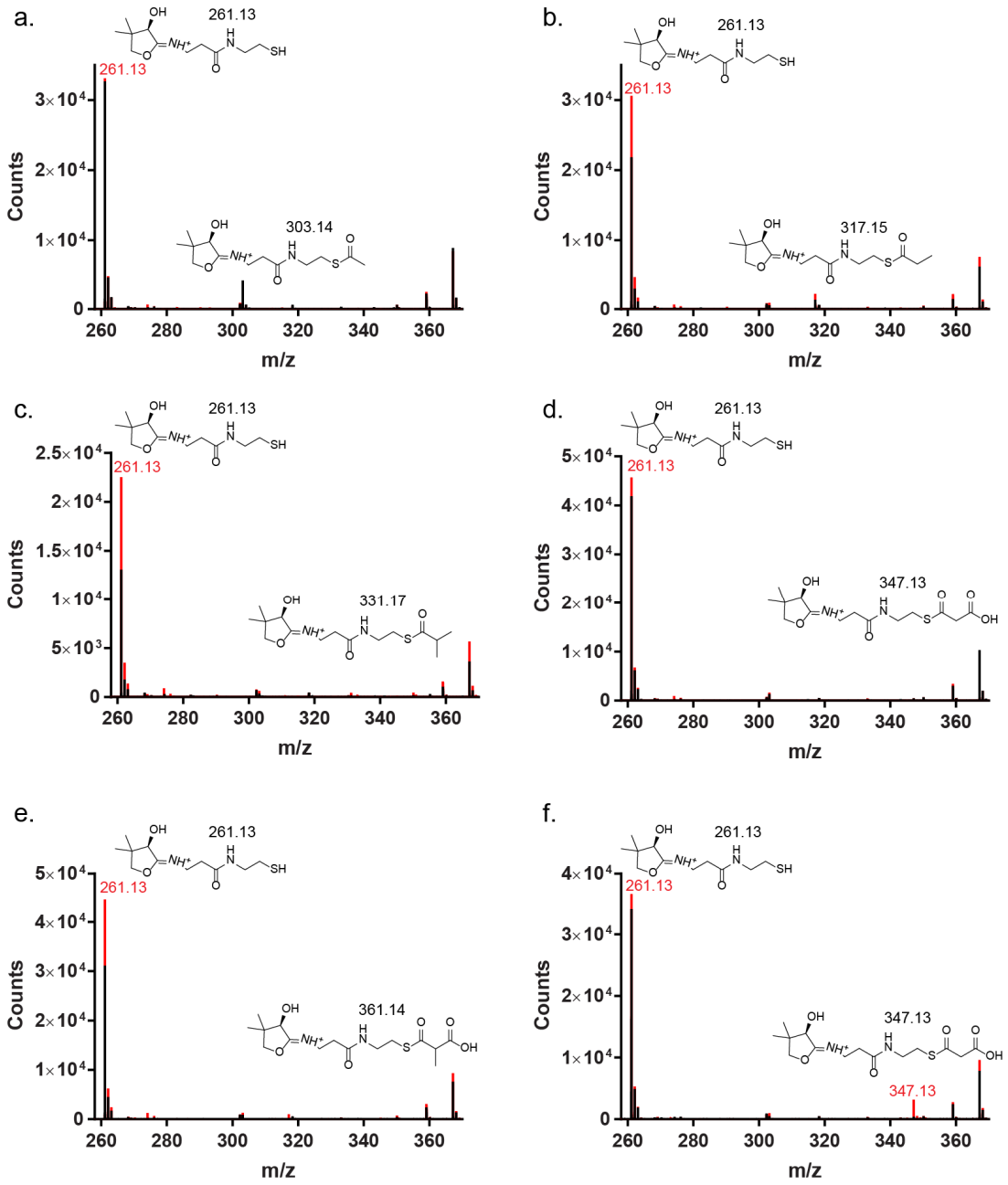


Figure 6.11 Representative mass spectra of GphF GNAT and MT-GNAT acyl transfer reactions

The 4-hr reaction mixtures included **a.** GphF GNAT, AprA holo-ACP, acetyl-CoA; **b.** GphF GNAT, AprA holo-ACP, propionyl-CoA; **c.** GphF GNAT, AprA holo-ACP, isobutyryl-CoA; **d.** GphF GNAT, AprA holo-ACP, Mal-CoA; **e.** GphF GNAT, AprA holo-ACP, MeMal-CoA; **f.** GphF MT_L-GNAT; AprA holo-ACP, Mal-CoA. AprA holo-ACP was used as a surrogate for insoluble GphF ACP_L. Spectra of no enzyme controls are shown in black; reactions are shown in red. Calculated m/z values are listed in black above structures of Ppant ejection fragments. Observed m/z values are indicated in red.

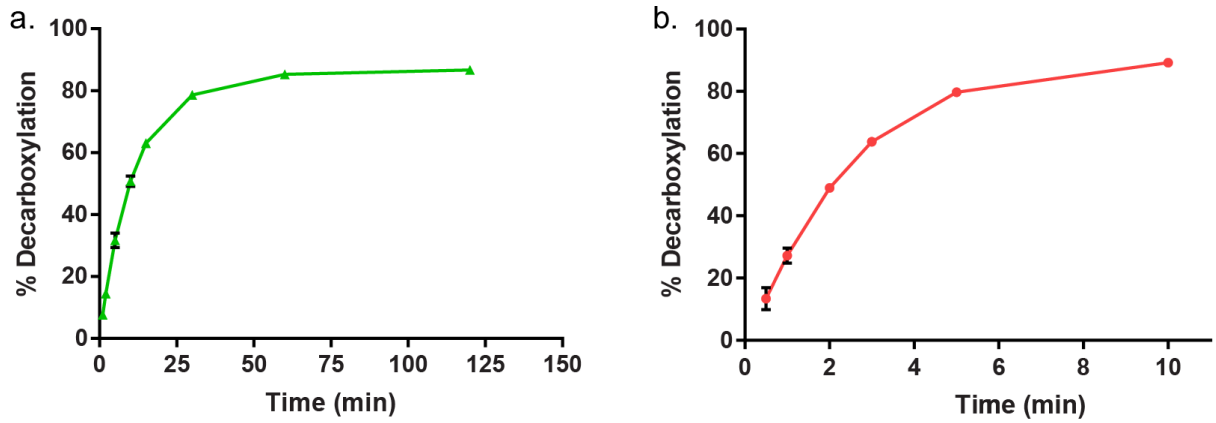


Figure 6.12 Decarboxylation by extensively purified GphF and CurA GNAT used for acyl transfer assays

Additional purification steps, which affect rates of acyl transfer do not alter rapid decarboxylation activity (compare with Figure. 6.5). **a.** Time course for Me_2Mal-ACP decarboxylation by GphF GNAT purified with Ni-affinity, size-exclusion, ion-exchange, and size-exclusion steps. **b.** Time course for Mal-ACP decarboxylation by CurA GNAT purified with Ni-affinity, TEV protease cleavage, Ni-affinity, and size-exclusion steps. Error bars represent triplicate experiments and, in some cases, are too small to be visible**.**

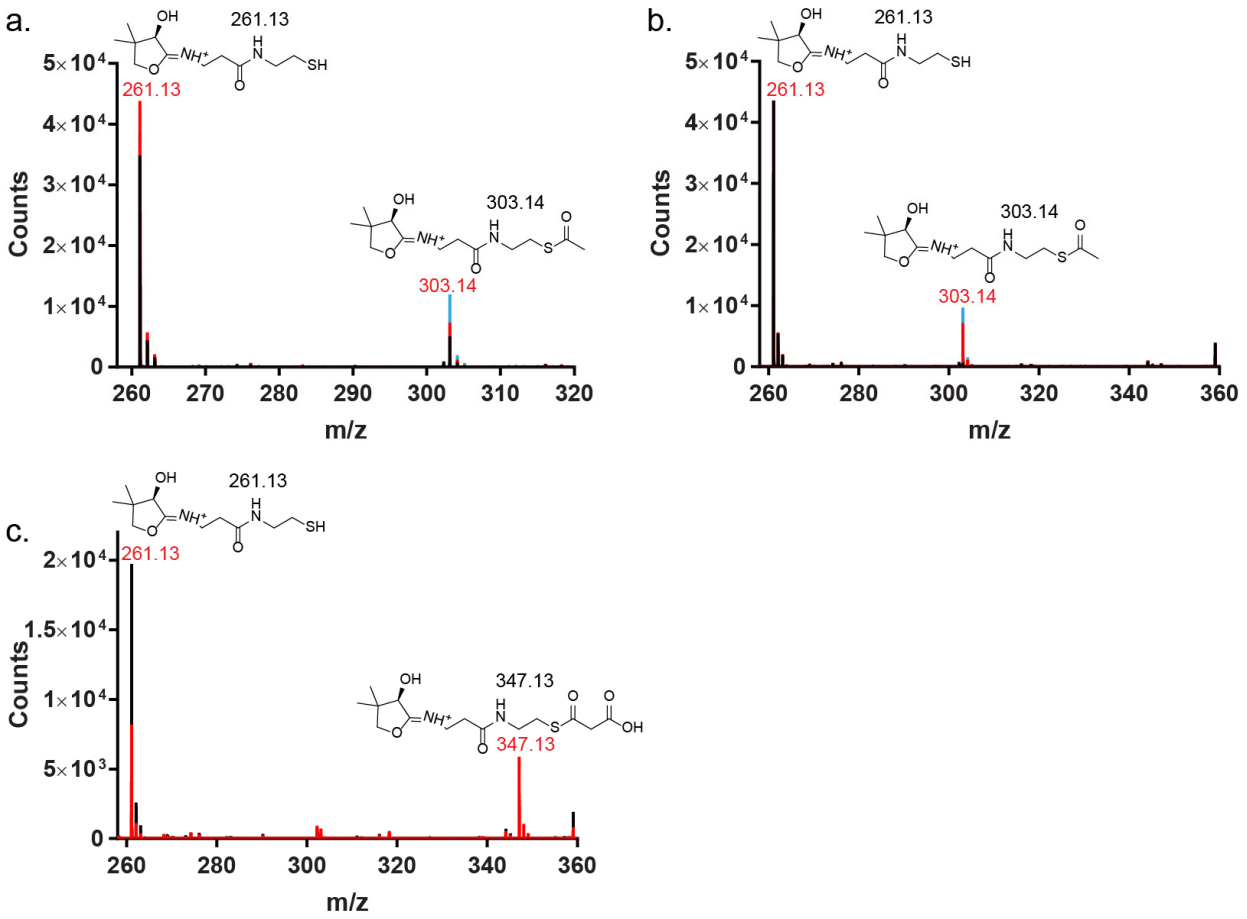


Figure 6.13 Representative mass spectra of CurA GNAT and *M. bouillonii* FabD acyl transfer reactions

The 4-hr reaction mixtures included **a.** CurA GNAT, CurA holo-ACP, acetyl-CoA; **b.** CurA GNAT, CurA holo-ACP, Mal-CoA. Spectra of no enzyme controls are shown in black; reactions with CurA GNAT purified with a single purification step are shown in blue; reactions with CurA GNAT purified with multiple purification steps are shown in red. All Mal-ACP in (b) was decarboxylated to acetyl-ACP by CurA GNAT. **c.** Malonyl transfer reaction catalyzed by *M. bouillonii* FabD, the malonyl acyltransferase of fatty acid biosynthesis. The 5-min reaction mixtures included CurA holo-ACP and Mal-CoA. Spectra of no enzyme controls are shown in black; reactions are shown in red. Calculated m/z values are listed in black above structures of Ppant ejection fragments. Observed m/z values are indicated in red.

PKS GNAT-like domains are acyl-ACP decarboxylases

Based on the results for the myxobacterial GphF $GNAT_L$ and the cyanobacterial CurA $GNAT_L$, decarboxylation is the primary role of GNAT-like enzymes in PKS initiation modules. As the rate of acyl transfer is non-existent or suspiciously low for these GNAT-like domains compared to their decarboxylation rates[31] and in some cases can be reduced further by additional purification steps, the PKS GNAT-like domain is more appropriately designated as an acyl-ACP decarboxylase. This places the GNAT-like domains of PKS initiation modules on the branch of the GNAT superfamily with the MCD decarboxylase[183, 242, 243]. These PKS acyl-ACP decarboxylases and MCD share the active site His and Ser/Thr residues essential for decarboxylation[183, 242].

A distant branch of the GNAT superfamily includes another enzyme of natural product biosynthesis, the malonyl- and methylmalonyl-CoA decarboxylase EryM (also known as SACE_1304 or Mcd[248-250]), which is encoded by *mcd* in *Saccharopolyspora erythrea* and is essential for the production of both the polyketide precursor of erythromycin[251] and the siderophore erythrochelin[249]. EryM/SACE_1034/Mcd is encoded in an otherwise defunct NRPS gene cluster separated from the erythromycin cluster by 0.65 Mbp and the erythrochelin cluster by ~2 Mbp. Disruption of the gene eliminates both erythromycin and erythrochelin production[249, 251]. The multifunctional EryM decarboxylates methylmalonyl-CoA to generate a propionyl-CoA starter unit for erythromycin biosynthesis[251]. The same protein catalyzes dual decarboxylation of malonyl-CoA and acetyl transfer to δ-*N*-hydroxy-L-ornithine in erythrochelin biosynthesis[249, 250]. Although the sequence identity is less than 20%, the C-terminal domain of EryM was recognized as a member of the GNAT superfamily[249, 250]. The GNAT superfamily membership is clear from the significant sequence identity (36%) to superfamily member *Mycobacterium tuberculosis* Rv1347c, which was itself recognized as a superfamily member from the crystal structure was solved[252]. The GNAT-like domain of EryM does not contain His and Ser/Thr residues in analogous locations in the GNAT fold to acyl-ACP decarboxylases and MCD. Instead EryM contains a His Glu/Asp dyad, which is conserved in sequences of unannotated homologs in bacterial genome sequences (Figure 6.14). Interestingly, the His Glu/Asp dyad is critical for transferring long chain fatty acids from an ACP to the primary amine of a siderophore in *M. tuberculosis* Rv1347c [253, 254] and likely plays a role in activating the acyl accepting amine of δ-*N*-hydroxy-L-ornithine in erythrochelin biosynthesis.

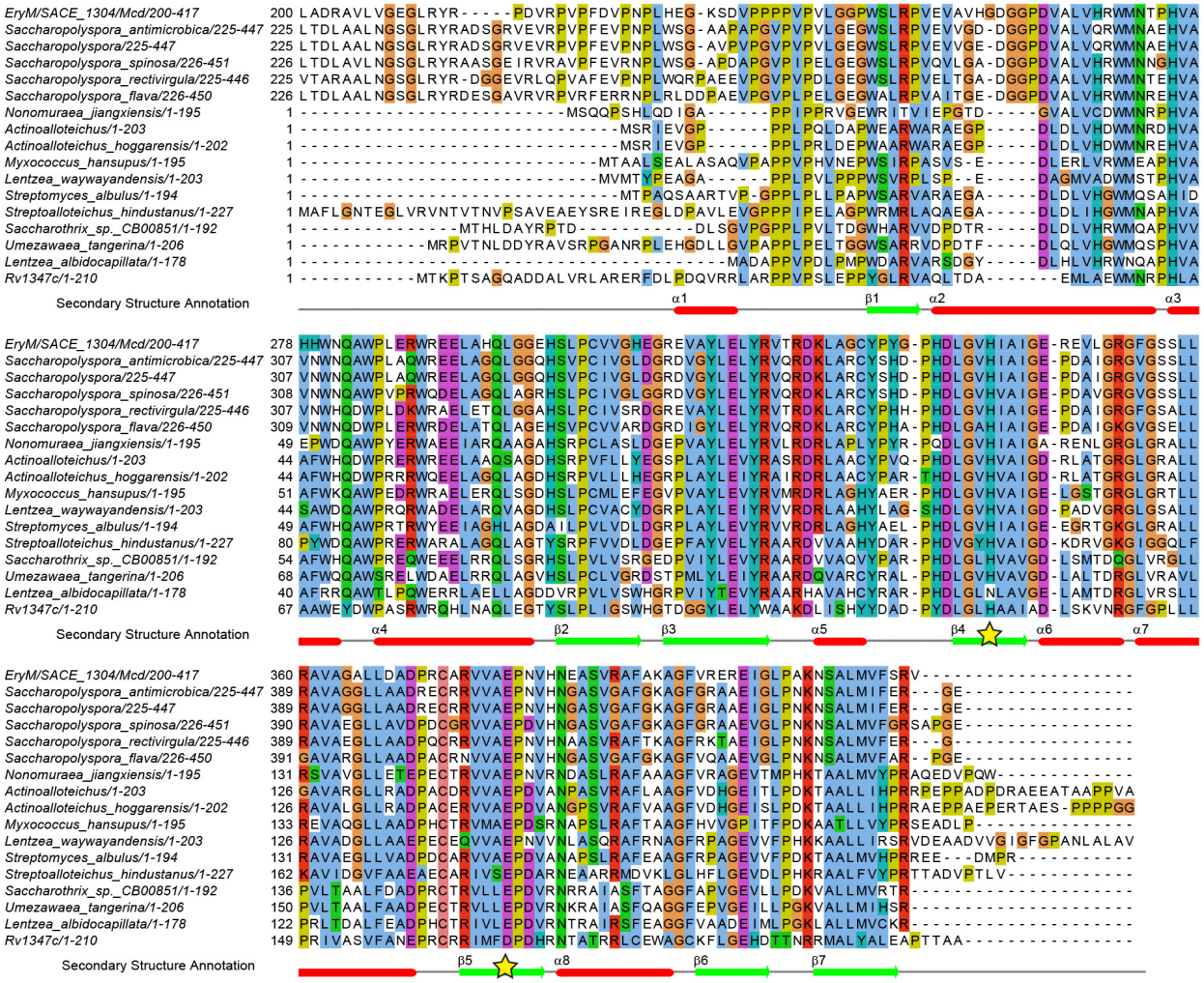


Figure 6.14 Sequence alignment of the GNAT domain of EryM / SACE_1304 / Mcd homologs and Rv1347c.

The Rv1347c secondary structure annotation is displayed below the alignment. Rv1347c catalytic residues, which are conserved in EryM homologs are starred. GenBank accession codes are as follows: EryM / SACE_1304 / Mcd (*Saccharopolyspora erythrea*), CAM00627.1; *Saccharopolyspora antimicrobica,* WP_093149902.1; *Saccharopolyspora*, WP_093351295.1; *Saccharopolyspora spinosa*, PKW14738.1; *Saccharopolyspora rectivirgula*, WP_051387681.1; *Saccharopolyspora flava*, SFT09105.1; *Nonomuraea jiangxiensis*, WP_090929840.1; *Actinoalloteichus*, WP_075740024.1; *Actinoalloteichus hoggarensis*, WP_093941036.1; *Myxococcus hansupus*, WP_002639252.1; *Lentzea waywayandensis*, WP_093605558.1; *Streptomyces albulus*, WP_064068670.1; *Streptoalloteichus hindustanus*, SHE93970.1; *Saccharothrix sp. CB00851*, WP_083668623.1; *Umezawaea tangerine*, PRY28657.1; *Lentzea albidocapillata*, WP_030481780.1; Rv1347c (*Mycobacterium tuberculosis*), WP_003406956.1.

Overall, we propose that the ACP_L of PKS initiation modules containing acyl-ACP decarboxylases is primed with malonyl by a stand-alone malonyl-acyltransferase not encoded in the PKS gene cluster. FabD, the malonyl-acyltransferase from fatty acid biosynthesis, supports rapid malonyl acyl transfer to AprA ACP_L[120] and CurA ACP_L, as well as other PKS ACPs[196, 233] We suspect that the extremely slow and variable malonyl acyltransferase activity of the MT_L or acyl-ACP decarboxylase (formerly known as a GNAT-like domain) is due to a contaminating *E. coli* enzyme, as additional purification steps of the acyl-ACP decarboxylases decreased acyl transfer activity but did not alter the rate of decarboxylation. Previously, the identification of contaminating activity from *E. coli* enzymes refuted a "self-loading" ability of PKS ACPs[196]. After malonyl acyl transfer, initiation modules containing no MT_L, such as CurA (Figure 6.1a) proceed directly to decarboxylation by the acyl-ACP decarboxylase (Figure 6.1b). Decarboxylation requires conserved His and Thr/Ser residues. We further demonstrated the importance of a conserved active site Arg in decarboxylation for both cyanobacterial and myxobacterial acyl-ACP decarboxylases.

Initiation modules with an MT_L, such as GphF (Figure 6.1c), undergo metal-dependent methylation to produce MeMal-ACP or Me_2Mal-ACP (Figure 6.1d) [119]. Methylation of an ACP-tethered malonyl rather than Mal-CoA is essential as MeMal and Me_2Mal are precious resources, not normally produced at sufficient quantities in proteobacteria and cyanobacteria[246] and must be sequestered by the PKS enzyme assembly line. Once the pathway-specific degree of methylation has occurred, the acyl-ACP decarboxylase will perform decarboxylation, and the resulting starter unit can be passed to the next module or catalytic domain for further processing (Figure 6.1d). The high degree of substrate selectivity of the acyl-ACP decarboxylase limits the production of the wrong starter unit, which may be incompatible with downstream pathway processing and may stall productive catalysis by the PKS assembly line. The crystal structures and mutagenesis studies of GphF and CurA acyl-ACP decarboxylases did not reveal the basis for the observed acyl group selectivity.

PKS acyl-ACP decarboxylases join MCD as GNAT superfamily members that catalyze decarboxylation only. At least one other decarboxylase branch of the GNAT superfamily exists and includes the acyltransferase Rv1347c[252] and the bifunctional decarboxylase / acyltransferase EryM/SACE_1034/Mcd essential for erythromycin and erythrochelin biosynthesis[251]. Additional decarboxylating members of the GNAT superfamily may be misannotated as acyltransferases.

The common function of the GNAT fold in the decarboxylating and acylating superfamily members is to bind the Ppant of CoA- or ACP-linked substrates and to employ an enolate intermediate in catalysis. The common ancestor of the GNAT superfamily, including the histone acetyltransferase (HAT) branch of GNATs involved in epigenetic regulation and the aminoglycoside *N*-acetyltransferases and PKS acyl-ACP decarboxylases of secondary metabolism, may have been MCD, an essential enzyme of primary metabolism.

Chapter 7 Conclusions and Future Directions

Overview

Presented here is the full biochemical and structural characterization of the four types of MTs currently annotated in modular PKS: extension module *C*-MTs and *O*-MTs and initiation module MT_L and $MT2_L$ domains. Characterization of MT_L activity facilitated the full functional annotation of a family of PKS initiation modules containing acyl-ACP decarboxylase (formally known as GNAT-like) domains. The study of MT_L and $MT2_L$ from the apratoxin A biosynthetic pathway revealed a new route toward an unusual *t*-butyl functional group in natural product biosynthesis.

Structure and Activity of PKS MT Domains

All four types of PKS MTs are members of the class I MT superfamily. The PKS *C*-MT and *O*-MT domains evolved from different branches of class I MTs, whereas MT_L and $MT2_L$ share a common ancestor with *C*-MTs. The four PKS MTs perform distinctly different reactions; *C*-MTs, *O*-MTs and MT_Ls use different catalytic machinery for methylation of three distinct substrates, whereas $MT2_L$ has adapted the *C*-MT catalytic machinery for coupled decarboxylation and methylation reactions.

α-methylation by C-MT domains

Characterization of the *C*-MT from the curacin A biosynthetic pathway[23] clearly demonstrated that PKS *C*-MTs methylate the α-carbon of the β-keto intermediate produced by the KS extension reaction and not Mal-ACP or the β-hydroxy intermediate produced by the KR[117]. Methylation of the β-keto intermediate is consistent with studies of bacterial *C*-MTs from the gephyronic acid biosynthetic pathway[103] and *trans*-AT pathways[102]. The crystal structure of CurJ *C*-MT contains a long hydrophobic substrate binding tunnel that can fully accommodate the β-keto polyketide substrate[117]. An analogous tunnel exists in a structure of a fungal type I PKS *C*-MT[165]. Methylation by CurJ *C*-MT requires a conserved His and Glu catalytic dyad, which likely deprotonates the substrate carbon to form the carbanion nucleophile. The stereochemical outcome

of the *C*-MT reaction could not be predicted from the CurJ structure, as the substrate-free active site is not fully set up for catalysis and the acidity of the stereogenic proton of the methylated product impedes chemoenzymatic assignment. Recently, the stereochemistry of the *C*-MT reaction was elucidated through a coupled MT-KR assay evaluated by chiral gas chromatography-mass spectrometry (GC-MS)[155].

A striking structural similarity between the CurJ *C*-MT and vestiges of an MT in the mFAS ΨMT domain, further solidifies the evolutionary relationship between mFAS and PKS. Analysis of the linker regions surrounding the CurJ *C*-MT facilitated modeling of interactions between the PKS *C*-MT and KR domains. In the model, the core of the *C*-MT extends from the KRs domain placing it in a similar position to the mFAS ΨMT. Such positioning supports the monomeric state of the excised *C*-MT and would readily permit the evolutionary gain or loss of a module's *C*-MT domain without disrupting the central PKS module architecture.

β-hydroxy methylation by O-MT domains

PKS *O*-MTs evolved from a different branch of class I MTs than *C*-MTs, with the *O*-MT structures closely resembling MTs from other secondary metabolite pathways. Crystal structures of the StiD and StiE *O*-MT from the stigmatellin biosynthetic pathway[19] defined the *O*-MT domain boundaries and revealed a Glu that is essential for methylation of the β-hydroxy intermediate by the myxobacterial StiE *O*-MT and cyanobacterial CurL *O*-MT from the curacin biosynthetic pathway[23]. The Glu is stabilized by a conserved Tyr and may facilitate deprotonation of the β-hydroxyl. Analysis of a diastereomeric substrate mimic library demonstrated that the *O*-MT reaction is stereospecific. Although the StiD and StiE structures represent *O*-MTs that methylate (*S*)- and (*R*)-β-hydroxyl groups, no structural determinant of stereospecificity was identified due to solvent exposed active sites. Visualization of a partially ordered *N*-terminal helix in the StiE *O*-MT structure suggests that the crystal structures do not capture the *O*-MT active site in a full catalytically competent state. Upon binding of the ACP linked substrate, an N-terminal helix, which is present in structures of homologs from other secondary metabolite pathways and is partially ordered in a structure of StiE *O*-MT, may order above the active site. The N-terminal helix is critical for activity, as truncation of the helix results in an inactive *O*-MT. The oligomeric state of excised PKS *O*-MTs is variable. The StiD and StiE structures indicate that *O*-MTs from myxobacteria are dimeric, whereas the cyanobacterial CurL *O*-MT is monomeric. The CurL module contains an additional dimerization domain N-terminal to the *O*-MT. As *O*-methylation

precludes the inclusion of reductive DH and ER domains in a module, the presence of an *O*-MT may stabilize the dimeric PKS module in lieu of a dimeric DH.

Malonyl methylation by MT_L domains

Found in GNAT containing PKS initiation modules, such as AprA, the structure of the MT_L domain most closely resembles PKS extension module *C*-MT domains despite sharing only 13% sequence identity[119]. The large AprA MT_L lid domain, which is also found independently in initiation modules lacking the MT_L core and was previously referred to as an "adaptor region" (AR) of unknown function, contains the full lid seen in PKS *C*-MTs with five additional helices appended. Thus, MT_L and *C*-MTs likely evolved from a common ancestor. However, MT_L adopted a completely different catalytic strategy to methylate the α-position of Mal-ACP, whereas *C*-MTs cannot methylate carboxylated substrates. Surprisingly, MT_L is a mononuclear-iron-dependent MT, joining only one other reported in the literature[74].

A malonate-bound AprA MT_L structure showed that the metal center coordinates two Mal-ACP carboxyl groups and acts as a Lewis acid to lower the pKa of the α-carbon and promote proton abstraction, enolate formation, and attack of the SAM methyl by the carbanion. In the presence of Fe^{3+}, AprA MT_L is capable of dimethylating Mal-ACP to Me_2Mal-ACP, whereas Ni^{2+}, Mn^{2+}, Fe^{2+}, and Co^{2+} exclusively support one methylation yielding MeMal-ACP. Additionally, a AprA active site Tyr to Phe variant supports only a single methylation. We hypothesize that removal of the remaining α-proton of MeMal-ACP, which has a higher pKa than a Mal-ACP α-proton, requires the stronger Lewis acid Fe^{3+}. The active site Tyr is positioned perfectly for the removal the α(*R*)-proton from the α(S)-methylmalonyl product of the first methylation reaction. The identity of the catalytic base to remove the α(S)-proton is unknown. Several amino acid substitutions in an elegant malonate-stabilizing hydrogen bond network eliminated methylation activity, but these substrate-stabilizing residues are not positioned correctly for proton abstraction. AprA MT_L has specifically evolved to promote dimethylation. The substrate can be threaded in through an active site entrance, which contains conserved amino acids that may stabilize the Mal-ACP Ppant. A cleft on the opposite side of the active site near the SAM binding site facilitates exchange of SAH for fresh SAM without disengaging the substrate from the metal center. The action of MT_L provides a new mechanism to generate MeMal-ACP and Me_2Mal-ACP for secondary metabolite biosynthesis in organisms that do not produce MeMal-CoA[246]. Finally,

characterization of MT_L facilitates the full functional annotation of many members of the GNAT-containing PKS initiation module family.

t-Butyl production by $MT2_L$

The $MT2_L$ domain from the AprA initiation module shares ~30% sequence identity with extension module *C*-MTs including the His-Glu catalytic dyad. In $MT2_L$ the *C*-MT fold has been adapted to catalyze decarboxylation and methylation reactions, directly transforming Me_2Mal-ACP to pivaloyl-ACP[120]. The two reactions are closely coupled, as variation of the degree of substrate methylation (e.g. Mal-ACP, MeMal-ACP, Me_2Mal-ACP) and many amino acid substitutions decoupled methylation from decarboxylation, yielding shunt products of decarboxylation only. However, only substitutions to the His-Glu dyad, as well as a drastic Gly to Phe variant, severely limited both methylation and decarboxylation reactions. Synthesis of *t*-butyl in the form of pivaloyl-ACP by $MT2_L$ drastically differs from the only other reported biosynthetic route to *t*-butyl in which a cobalamin-dependent radical SAM enzyme methylates valine derivatives[202, 203]. Functional annotation of AprA provides a "barcode" for genome mining of *t*-butylated natural products. For years, the function of BryX in the biosynthesis of the protein kinase C modulator bryostatin was unknown[20]. Reannotation of the bryostatin pathway identified BryX as an AprA homolog. Biochemical analysis confirmed that BryX $MT2_L$ is also capable of producing pivaloyl-ACP, classifying BryX as the module responsible for *t*-butyl production in bryostatin biosynthesis.

Functional Annotation of GNAT Containing Initiation Modules

Discovery of the ΨGNAT domain

Structures of AprA MT_L-GNAT and GNAT-$MT2_L$ revealed that the AprA GNAT domain is truncated and lacks residues essential for catalysis in active GNAT domains[31, 219]. Decarboxylation and acyl transfer activities, which were expected for PKS GNAT-like domains, were not supported by AprA GNAT. Therefore, the AprA and homologous BryX GNAT were reclassified as vestigial "pseudo"-GNATs (ΨGNATs). The ΨGNAT plays a structural role in the architecture of the initiation module, as MT_L domains cannot be solubly produced without the neighboring GNAT domain. The remaining AprA domains, MT_L and $MT2_L$, did not support acyl transfer and no acyltransferases or domains without assigned function are encoded in the apratoxin

A gene cluster[25]. Thus, an enzyme encoded outside the gene cluster must provide the missing acyl transfer step, most likely priming the AprA ACP_L with malonyl. FabD, the malonyl acyltransferase from the apratoxin A producer *Moorea bouillonii*, is an essential enzyme that provides the malonyl acyl transfer step from Mal-CoA to the ACP (AcpP) in bacterial fatty acid biosynthesis. *M. bouillonii* FabD supports rapid malonyl acyltransfer to AprA ACP and may be the acyltransferase responsible for initiating apratoxin A biosynthesis, linking primary and secondary metabolism in *M. bouillonii*.

Decarboxylation activity

The discovery of the production of MeMal-ACP and Me_2Mal-ACP by MT_L prompted further investigation of the GNAT family of initiation modules. The GNAT-like domain was hypothesized to perform malonyl acyl transfer from CoA to ACP_L prior to MT_L methylation. Once methylation occurs, ACP_L was thought to return to the GNAT-like domain for decarboxylation yielding the starter unit for the biosynthetic pathway. To better understand the reaction sequence of the GNAT family of loading modules, we investigated the GphF GNAT, which produces an isobutyryl starter unit in the gephyronic acid biosynthetic pathway[21]. GphF displayed strong substrate selectivity for Me_2Mal-ACP over MeMal-ACP and no activity on Mal-ACP. Similarly, CurA GNAT, the GNAT-like domain from the curacin A biosynthetic pathway[23], preferred to decarboxylate Mal-ACP over MeMal-ACP and Me_2Mal-ACP, correlating with the production of an acetyl starter unit. Therefore, the GNAT-like domain acts as a gatekeeper, selecting decarboxylation substrates with the proper degree of methyaltion.

A crystal structure of the GphF GNAT in complex with the product mimic isobuturyl-CoA provided additional insight into GNAT reactivity. Mutagenesis studies of GphF and CurA GNAT did not identify specific residues that influence the substrate selectivity of PKS GNAT-like domains. However, a conserved active site Arg key for decarboxylation was identified. The location of the active site Arg and modeling of the carboxylated substrate based on the isobutyryl bound GphF structure indicate that the Arg may coordinate the terminal carboxylate of the ACP bound substrate. The conserved active site His and Ser/Thr residues required for decarboxylation likely stabilize the enolate intermediate and re-protonate the carbanion upon collapse of the enolate.

Acyl transfer activity

The rate of GphF GNAT acyltransfer varied between enzyme preparations, whereas the rate of decarboxylation was relatively constant. The fluctuating rates of activity suggested that different batches of GphF GNAT could contain varying levels of a contaminating acyltransferase and prompted further purification of GphF GNAT. Stringently purified GphF GNAT did not catalyze acyl transfer between a panel of acyl-CoA substrates and an ACP. Additionally, further purification of CurA GNAT diminished the slow rate of acetyl transfer, but not decarboxylation. The fluctuating rates of acyl transfer activity for PKS GNAT-like domains was attributed to a contaminating acyltransferase from heterologous expression in *E. coli*. Precedent for such activity comes from a reported "self-loading" activity by PKS ACP domains, which was later credited to a contaminating *E. coli* malonyl acyltransferase[196]. Caution must be taken when evaluating enzymatic reactions using simple acyl-CoA substrates, which are substrates for highly active primary metabolic enzymes. In fact, the "self-loading" ACP is not the only example of reported activity in PKS systems that was later attributed to a contaminant; in erythromycin biosynthesis a KS was implicated in the decarboxylation of methylmalonyl-CoA to propionyl-CoA, such activity was not detected with rigorously purified protein and was concluded to be a result of a contaminating decarboxylase[255].

The lack of acyl transfer activity for PKS GNAT-like enzymes prompted reannotation of the domains as acyl-ACP decarboxylases, which reflects their function. Such nomenclature is consistent with the primary metabolism enzyme malonyl-CoA decarboxylase[183, 242] and EryM/SACE_1304/Mcd[249-251], a bifunctional decarboxylase / acetyltransferase, both of which are also decarboxylating members of the GNAT superfamily. Still, the malonyl acyl transfer step was unaccounted for in the acyl-ACP decarboxylase family of initiation modules. As in apratoxin A biosynthesis, the bacterial FAS malonyl acyltransferase, FabD, supported transfer of malonyl from CoA to CurA ACP_L. Thus, FabD is an attractive candidate for initiating biosynthesis in pathways with acyl-ACP decarboxylase-containing initiation modules. The malonyl acyl transfer step is critical for pathway throughput, as CoA bound substrates, which are not compatible with downstream pathway extension modules, could readily diffuse away from the biosynthetic assembly line, reducing its efficiency.

Architecture of a PKS Initiation Module

Structures of AprA MT_L-ΨGNAT and ΨGNAT-$MT2_L$ were combined to create a composite model of the full length AprA initiation module, which was tested using negative stain electron microscopy (EM). The EM micrographs displayed particles in a linear state resembling the model, but also particles in a bent state and a variety of intermediate conformations. The linear and bent states may be associated with MT_L and $MT2_L$ catalytic steps. A low-resolution EM envelope enabled docking of MT_L-ΨGNAT and ΨGNAT-$MT2_L$ crystal structures into the bent state. The ΨGNAT domain acts as a hinge, allowing MT_L to swing relative to the central $MT2_L$ dimer and access both the linear and bent states. The EM data indicates that PKS modules can be more dynamic than PikAIII[53, 60], the only structurally characterized full length PKS module, which displayed a relatively stable arched architecture.

Future Directions

Identification of novel natural products gene clusters

Implication of the mysterious BryX module as the machinery responsible for producing the *t*-butyl moiety in bryostatin bioynthesis clearly demonstrates the utility of the AprA sequence as a barcode for the identification of *t*-butylated natural products in bacterial genomes. In recent months, several genomes from a variety of bacterial sources containing an AprA/BryX like module have been deposited[256-259], indicating that the *$MT2_L$* strategy for the production of *t*-butyl groups may occur in many branches of the bacterial world. As the t-butyl group is a relatively uncommon feature in natural products, these organisms likely produce novel secondary metabolites.

Engineering new polyketides and the development of biocatalysts

Insight into the activity of PKS *C*-MT, *O*-MT, MT_L and $MT2_L$ domains has implications for the engineering of new polyketides, as well as the development of biocatalysts for synthetically challenging regio- and stereospecific methylation. Simple mutagenesis of the newly identified active site residues in *C*-MTs and *O*-MTs can produce novel demethylated polyketides, which may have altered bioactivities. This strategy has already been implemented for a *C*-MT involved in the biosynthesis of neoantimycin, a commonly used piscicide, to create a demethylated neoantimycin derivative through CRISPR/Cas9 genome editing of the neoantimycin gene cluster in a heterologous expression strain[260].

Information on substrate selectively is limited for all PKS domains due to a lack of substrate-bound crystal structures. Thus, efforts to engineer *C*-MTs and *O*-MTs to methylate additional substrates is restricted to directed evolution strategies, as the features of the MT domains that determine substrate selectivity are unknown. The low affinity between ACP bound substrates and PKS enzymatic domains and the limited stability of the thioester linkage between the Ppant and acyl substrate precludes the attainment of substrate-bound structures by x-ray crystallography. Several synthetic approaches could yield a substrate mimic to facilitate the determination of substrate-bound *C*-MT or *O*-MT structures. Polyketide substrates or close substrate analogs could be appended to a nonhydrolyzable phosphopantetheine moiety. The stable substrate and inclusion of the full Ppant, compared to traditionally used *N*-acetylcysteamine Ppant mimic, may provide enhanced affinity for the MT domain in co-crystallization experiments. Such an approach was successfully used to obtain a substrate-bound type II PKS KS crystal structure[261]. Alternatively, co-crystallization of a *C*-MT or *O*-MT with a substrate-linked SAM analog could also provide valuable information on substrate positioning in the active site. Such probes are similar to those developed as MT inhibitors[262]. Capturing a *C*-MT with a truncated mimic of the β-keto substrate linked to SAM could provide insight into why some *C*-MTs methylate once, whereas others produce gem-dimethyl groups. Similarly, the development of a SAM linked β-hydroxy substrate mimic could hint at the structural determinants PKS *O*-MT stereospecificity.

Variants of MT_L found in different biosynthetic pathways can either mono- or dimethylate the malonyl group. A single amino acid substitution or provision of a metal other than Fe^{3+} transformed the dimethylating AprA MT_L into an MT capable of only one methylation. Further investigation of metal ligands and active site amino acids in other MT_Ls could provide additional understanding of features that influence the degree of MT_L methylation, leading to the production of new natural products with varying methyl groups through mutagenesis. Additionally, the MT_L metal center provides a scaffold for methylation of a wide degree of carboxylated substrates. While optimizing AprA MT_L crystals we observed clear density for carboxylated buffer components (e.g. citrate, tartrate) coordinated to the metal center, indicating that the MT_L active site can accommodate a variety of substrates other than malonyl or methylmalonyl. The buffer components did not appear to be methylated by SAM, which was included in the crystallization condition. Structural information could guide the placement of new catalytic machinery into the MT_L active

site to generate a biocatalytic route for stereo and regioselective methylation of substrates that are challenging to methylate synthetically.

Understanding interdomain interactions and PKS module architecture

High resolution structural information for PKS multi-domains and intact modules is essential for fully understanding PKS biosynthesis. Structures of the PKS *C*-MTs and *O*-MTs provided clues regarding the placement of MT domain within PKS modules. Crystal structures of a MT-KR didomain or a DH-MT-KR tridomain would provide valuable information regarding interdomain interactions and the role of linker regions in the positioning of the MT domains relative to other domains. Additional structures of intact PKS modules would be even more informative, as they would provide views of each domain in their truly native environment. Negative stain analysis of the AprA initiation module showed a high degree of flexibility of a module containing only three domains. PKS extension modules may possess similarly high degrees of flexibility, which could intensify with the inclusion of additional catalytic domains. The flexibility of modification domains is further supported by a recent crystal structure of a DH-KR-ER tri-domain, where domains rotate up to 40° between copies of the tri-domain in the crystal's asymmetric unit[3].

The success in characterizing the intact PKS module PikAIII by moderate resolution cryo-EM may be attributed to its relatively simple tridomain architecture and access to native substrates, which localized the ACP in defined chemical states[53, 60]. The inherent flexibility of PKS modules suggests that EM is a more suitable technique than crystallography for the determination of structures of modules with additional domains, such as MTs. Although analysis of interdomain interactions in PikAIII was restricted due to the resolution limitations of the structure, new EM methodology may provide higher resolution data for PKS modules[263]. However, strategies must be taken to trap the module in a homogenous population for EM, which will simplify the structure determination process and enhance resolution. As seen in PikAIII, the provision of native substrates can stabilize interactions with the ACP and possibly limit a module's flexibility. Therefore, the availability of synthetic substrates is essential for future PKS EM targets. If ACP localization is not sufficient, the development of conformationally specific nanobodies to specific chemical states of the PKS module could be explored[264]. Such an approach would likely require nonhydrolyzable ACP linked substrates and additional screening with nanobodies and excised domains to filter out nanobodies that bind only to single domains and may not stabilize the overall

architecture. Characterization of an intact PKS module containing all modification domains (KS-AT-DH-MT-KR-ER-ACP) is most imperative, as it would provide an answer to one of the most fundamental question regarding polyketide biosynthesis: do PKS modules resemble mFAS?

Evolutionary implications

The structural resemblance between the PKS *C*-MT and mFAS ΨMT further supports a common ancestor for PKS and mFAS and raises the possibility of an ancestral methylating FAS, as the existence of the metabolically essential mFAS likely preceded the rise of PKS secondary metabolite biosynthetic pathways. Genome sequences show that the FAS ΨMT is common to a diverse set of metazoans, including arthropods (insets), chordates (fish, amphibians, reptiles, birds), cnidaria (sea anemones, coral, jelly fish), echinoderms (starfish, sea urchins), mollusks (scallops, oysters), and even incredibly simple animals such as the placozoa *Trichoplax adhaerens*. Therefore, a FAS with the capability of producing methylated fatty acids likely predates metazoans.

A similar evolutionary conundrum exists for the enormous GNAT superfamily, which encompasses a wide range of enzymes including the acyl-ACP decarboxylases found in PKS initiation modules, the primary metabolic enzyme malonyl-CoA decarboxylase (MCD), histone *N*-acetyltransferases (HATs), and aminoglycoside *N*-acetyltransferases. GNATs are almost exclusively known for their acetyltransferase function. However, the superfamily may have evolved from an enzyme capable of binding the Ppant arm of CoA or ACP linked substrates, which is a common feature of both the GNAT acyltransferases and decarboxylases. Perhaps the primary metabolism enzyme MCD is actually the ancestral GNAT, from which all CoA binding members of the GNAT superfamily evolved. The discovery of decarboxylases currently misannotated as GNAT acyltransferases in the sequence databases and structural characterization of additional decarboxylating members of the GNAT family, such as EryM/SACE_1304/Mcd[251] the decarboxylating GNAT from the erythromycin producer *Saccharopolyspora erythraea*, which has no sequence identity to PKS acyl-ACP decarboxylases or malonyl-CoA decarboxylase, could provide further insight into the evolutionary origin of the GNAT fold.

High expression levels of the GNAT family member MCD occur in the uropygial gland of birds, where multimethyl-branched fatty acids that form a feather protecting oil are synthesized[244]. MCD reduces Mal-CoA levels in the uropygial gland, forcing mFAS to use available MeMal-CoA extender units for fatty acid biosynthesis, effectively producing the same product as mFAS with

an active *C*-MT domain. Perhaps the need for a *C*-MT in an ancestral mFAS diminished with the rise of MCD and the MeMal-CoA producing enzyme propionyl-CoA carboxylase, leaving the vestigial ΨMT observed in modern mFAS. Therefore, hints towards the evolutionary origin of mFAS and tangentially PKS may lie within the evolution of metabolic enzymes peripheral to fatty acid biosynthesis.

Bibliography

1. Maier, T., Leibundgut, M., and Ban, N. The crystal structure of a mammalian fatty acid synthase, *Science 321*, 1315-1322 (2008).
2. Keatinge-Clay, A. T., and Stroud, R. M. The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases, *Structure 14*, 737-748 (2006).
3. Herbst, D. A., Jakob, R. P., Zahringer, F., and Maier, T. Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases, *Nature 531*, 533-537 (2016).
4. Bonnett, S. A., Whicher, J. R., Papireddy, K., Florova, G., Smith, J. L., and Reynolds, K. A. Structural and stereochemical analysis of a modular polyketide synthase ketoreductase domain required for the generation of a cis-alkene, *Chem Biol 20*, 772-783 (2013).
5. Zheng, J., Gay, D. C., Demeler, B., White, M. A., and Keatinge-Clay, A. T. Divergence of multimodular polyketide synthases revealed by a didomain structure, *Nat Chem Biol 8*, 615-621 (2012).
6. Rossiter, S. E., Fletcher, M. H., and Wuest, W. M. Natural products as platforms to overcome antibiotic resistance, *Chem Rev 117*, 12415-12474 (2017).
7. Nagle, D. G., and Paul, V. J. Production of secondary metabolites by filamentous tropical marine cyanobacteria: ecological functions of the compounds, *J Phycol 35*, 1412-1421 (1999).
8. Perry, R. D., Balbo, P. B., Jones, H. A., Fetherston, J. D., and DeMoll, E. Yersiniabactin from *Yersinia pestis*: biochemical characterization of the siderophore and its role in iron transport and regulation, *Microbiology 145 (Pt 5)*, 1181-1190 (1999).
9. Williams, D. H., Stone, M. J., Hauck, P. R., and Rahman, S. K. Why are secondary metabolites (natural products) biosynthesized?, *J Nat Prod 52*, 1189-1208 (1989).
10. Ji, H. F., Li, X. J., and Zhang, H. Y. Natural products and drug discovery. Can thousands of years of ancient medical knowledge lead us to new and powerful drug combinations in the fight against cancer and dementia?, *EMBO Rep 10*, 194-200 (2009).
11. Newman, D. J., and Cragg, G. M. Natural products as sources of new drugs from 1981 to 2014, *J Nat Prod 79*, 629-661 (2016).
12. Li, J. W., and Vederas, J. C. Drug discovery and natural products: end of an era or an endless frontier?, *Science 325*, 161-165 (2009).
13. Wright, G. D. Opportunities for natural products in 21(st) century antibiotic discovery, *Nat Prod Rep 34*, 694-701 (2017).
14. Osbourn, A. Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation, *Trends Genet 26*, 449-457 (2010).
15. Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Wieland Brown, L. C., Mavrommatis, K., Pati, A., Godfrey, P. A., Koehrsen, M., Clardy, J., Birren, B. W., Takano, E., Sali, A., Linington, R. G., and Fischbach, M. A. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters, *Cell 158*, 412-421 (2014).

16. Fischbach, M. A., Walsh, C. T., and Clardy, J. The evolution of gene collectives: how natural selection drives chemical innovation, *Proc Natl Acad Sci U S A 105*, 4601-4608 (2008).
17. Tibrewal, N., and Tang, Y. Biocatalysts for natural product biosynthesis, *Annu Rev Chem Biomol Eng 5*, 347-366 (2014).
18. Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J., and Leadlay, P. F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*, *Nature 348*, 176-178 (1990).
19. Gaitatzis, N., Silakowski, B., Kunze, B., Nordsiek, G., Blocker, H., Hofle, G., and Muller, R. The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase, *J Biol Chem 277*, 13082-13090 (2002).
20. Sudek, S., Lopanik, N. B., Waggoner, L. E., Hildebrand, M., Anderson, C., Liu, H., Patel, A., Sherman, D. H., and Haygood, M. G. Identification of the putative bryostatin polyketide synthase gene cluster from "*Candidatus Endobugula sertula*", the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*, *J Nat Prod 70*, 67-74 (2007).
21. Young, J., Stevens, D. C., Carmichael, R., Tan, J., Rachid, S., Boddy, C. N., Muller, R., and Taylor, R. E. Elucidation of gephyronic acid biosynthetic pathway revealed unexpected SAM-dependent methylations, *J Nat Prod 76*, 2269-2276 (2013).
22. Bushley, K. E., Raja, R., Jaiswal, P., Cumbie, J. S., Nonogaki, M., Boyd, A. E., Owensby, C. A., Knaus, B. J., Elser, J., Miller, D., Di, Y., McPhail, K. L., and Spatafora, J. W. The genome of tolypocladium inflatum: evolution, organization, and expression of the cyclosporin biosynthetic gene cluster, *PLoS Genet 9*, e1003496 (2013).
23. Chang, Z., Sitachitta, N., Rossi, J. V., Roberts, M. A., Flatt, P. M., Jia, J., Sherman, D. H., and Gerwick, W. H. Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*, *J Nat Prod 67*, 1356-1367 (2004).
24. Edwards, D. J., Marquez, B. L., Nogle, L. M., McPhail, K., Goeger, D. E., Roberts, M. A., and Gerwick, W. H. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*, *Chem Biol 11*, 817-833 (2004).
25. Grindberg, R. V., Ishoey, T., Brinza, D., Esquenazi, E., Coates, R. C., Liu, W. T., Gerwick, L., Dorrestein, P. C., Pevzner, P., Lasken, R., and Gerwick, W. H. Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage, *PLoS One 6*, e18565 (2011).
26. Huang, K. C., Chen, Z., Jiang, Y., Akare, S., Kolber-Simonds, D., Condon, K., Agoulnik, S., Tendyke, K., Shen, Y., Wu, K. M., Mathieu, S., Choi, H. W., Zhu, X., Shimizu, H., Kotake, Y., Gerwick, W. H., Uenaka, T., Woodall-Jappe, M., and Nomoto, K. Apratoxin A shows novel pancreas-targeting activity through the binding of Sec 61, *Mol Cancer Ther 15*, 1208-1216 (2016).
27. Paatero, A. O., Kellosalo, J., Dunyak, B. M., Almaliti, J., Gestwicki, J. E., Gerwick, W. H., Taunton, J., and Paavilainen, V. O. Apratoxin kills cells by direct blockade of the Sec61 protein translocation channel, *Cell Chem Biol 23*, 561-566 (2016).
28. Keatinge-Clay, A. T. The structures of type I polyketide synthases, *Nat Prod Rep 29*, 1050–1073 (2012).
29. Hertweck, C., Luzhetskyy, A., Rebets, Y., and Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork, *Nat Prod Rep 24*, 162-190 (2007).

30. Shimizu, Y., Ogata, H., and Goto, S. Type III polyketide synthases: functional classification and phylogenomics, *Chembiochem 18*, 50-65 (2017).
31. Gu, L., Geders, T. W., Wang, B., Gerwick, W. H., Hakansson, K., Smith, J. L., and Sherman, D. H. GNAT-like strategy for polyketide chain initiation, *Science 318*, 970-974 (2007).
32. Gu, L., Wang, B., Kulkarni, A., Geders, T. W., Grindberg, R. V., Gerwick, L., Hakansson, K., Wipf, P., Smith, J. L., Gerwick, W. H., and Sherman, D. H. Metamorphic enzyme assembly in polyketide diversification, *Nature 459*, 731-735 (2009).
33. Maloney, F. P., Gerwick, L., Gerwick, W. H., Sherman, D. H., and Smith, J. L. Anatomy of the β-branching enzyme of polyketide biosynthesis and its interaction with an acyl-ACP substrate, *Proc Natl Acad Sci U S A 113*, 10316-10321 (2016).
34. Khare, D., Wang, B., Gu, L., Razelun, J., Sherman, D. H., Gerwick, W. H., Hakansson, K., and Smith, J. L. Conformational switch triggered by alpha-ketoglutarate in a halogenase of curacin A biosynthesis, *Proc Natl Acad Sci U S A 107*, 14099-14104 (2010).
35. Geders, T. W., Gu, L., Mowers, J. C., Liu, H., Gerwick, W. H., Hakansson, K., Sherman, D. H., and Smith, J. L. Crystal structure of the ECH2 catalytic domain of CurF from *Lyngbya majuscula*. Insights into a decarboxylase involved in polyketide chain beta-branching, *J Biol Chem 282*, 35954-35963 (2007).
36. Khare, D., Hale, W. A., Tripathi, A., Gu, L., Sherman, D. H., Gerwick, W. H., Hakansson, K., and Smith, J. L. Structural basis for cyclopropanation by a unique enoyl-acyl carrier protein reductase, *Structure 23*, 2213-2223 (2015).
37. Gu, L., Eisman, E. B., Dutta, S., Franzmann, T. M., Walter, S., Gerwick, W. H., Skiniotis, G., and Sherman, D. H. Tandem acyl carrier proteins in the curacin biosynthetic pathway promote consecutive multienzyme reactions with a synergistic effect, *Angew Chem Int Ed Engl 50*, 2795-2798 (2011).
38. Gu, L., Wang, B., Kulkarni, A., Gehret, J. J., Lloyd, K. R., Gerwick, L., Gerwick, W. H., Wipf, P., Hakansson, K., Smith, J. L., and Sherman, D. H. Polyketide decarboxylative chain termination preceded by *O*-sulfonation in curacin a biosynthesis, *J Am Chem Soc 131*, 16033-16035 (2009).
39. Gehret, J. J., Gu, L., Gerwick, W. H., Wipf, P., Sherman, D. H., and Smith, J. L. Terminal alkene formation by the thioesterase of curacin A biosynthesis: structure of a decarboxylating thioesterase, *J Biol Chem 286*, 14445-14454 (2011).
40. McCarthy, J. G., Eisman, E. B., Kulkarni, S., Gerwick, L., Gerwick, W. H., Wipf, P., Sherman, D. H., and Smith, J. L. Structural basis of functional group activation by sulfotransferases in complex metabolic pathways, *ACS Chem Biol 7*, 1994-2003 (2012).
41. Keatinge-Clay, A. T. The Uncommon Enzymology of Cis-Acyltransferase Assembly Lines, *Chem Rev 117*, 5334-5366 (2017).
42. Salah Ud-Din, A. I., Tikhomirova, A., and Roujeinikova, A. Structure and functional diversity of GCN5-Related *N*-acetyltransferases (GNAT), *Int J Mol Sci 17* (2016).
43. Favrot, L., Blanchard, J. S., and Vergnolle, O. Bacterial GCN5-related *N*-acetyltransferases: from resistance to regulation, *Biochemistry 55*, 989-1002 (2016).
44. Piel, J., Wen, G., Platzer, M., and Hui, D. Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase, *Chembiochem 5*, 93-98 (2004).
45. Kellmann, R., Mihali, T. K., Jeon, Y. J., Pickford, R., Pomati, F., and Neilan, B. A. Biosynthetic intermediate analysis and functional homology reveal a saxitoxin gene cluster in cyanobacteria, *Appl Environ Microbiol 74*, 4044-4053 (2008).

46. Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F., and Weissman, K. J. The structure of docking domains in modular polyketide synthases, *Chemistry & Biology 10*, 723-731 (2003).
47. Buchholz, T. J., Geders, T. W., Bartley, F. E., 3rd, Reynolds, K. A., Smith, J. L., and Sherman, D. H. Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains, *ACS Chem Biol 4*, 41-52 (2009).
48. Whicher, J. R., Smaga, S. S., Hansen, D. A., Brown, W. C., Gerwick, W. H., Sherman, D. H., and Smith, J. L. Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis, *Chem Biol 20*, 1340-1351 (2013).
49. Fischbach, M. A., and Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms, *Chem Rev 106*, 3468-3496 (2006).
50. Tang, Y., Chen, A. Y., Kim, C. Y., Cane, D. E., and Khosla, C. Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase, *Chem Biol 14*, 931-943 (2007).
51. Tang, Y., Kim, C. Y., Mathews, II, Cane, D. E., and Khosla, C. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase, *Proc Natl Acad Sci U S A 103*, 11124-11129 (2006).
52. Zheng, J., Fage, C. D., Demeler, B., Hoffman, D. W., and Keatinge-Clay, A. T. The missing linker: a dimerization motif located within polyketide synthase modules, *ACS Chem Biol 8*, 1263-1270 (2013).
53. Dutta, S., Whicher, J. R., Hansen, D. A., Hale, W. A., Chemler, J. A., Congdon, G. R., Narayan, A. R., Hakansson, K., Sherman, D. H., Smith, J. L., and Skiniotis, G. Structure of a modular polyketide synthase, *Nature 510*, 512-517 (2014).
54. Keatinge-Clay, A. Crystal structure of the erythromycin polyketide synthase dehydratase, *J Mol Biol 384*, 941-953 (2008).
55. Akey, D. L., Razelun, J. R., Tehranisa, J., Sherman, D. H., Gerwick, W. H., and Smith, J. L. Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway, *Structure 18*, 94-105 (2010).
56. Akey, D. L., Kittendorf, J. D., Giraldes, J. W., Fecik, R. A., Sherman, D. H., and Smith, J. L. Structural basis for macrolactonization by the pikromycin thioesterase, *Nat Chem Biol 2*, 537-542 (2006).
57. Tsai, S. C., Miercke, L. J., Krucinski, J., Gokhale, R., Chen, J. C., Foster, P. G., Cane, D. E., Khosla, C., and Stroud, R. M. Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel, *Proc Natl Acad Sci U S A 98*, 14808-14813 (2001).
58. Keatinge-Clay, A. T. A tylosin ketoreductase reveals how chirality is determined in polyketides, *Chem Biol 14*, 898-908 (2007).
59. Smith, S., and Tsai, S. C. The type I fatty acid and polyketide synthases: a tale of two megasynthases, *Nat Prod Rep 24*, 1041-1072 (2007).
60. Whicher, J. R., Dutta, S., Hansen, D. A., Hale, W. A., Chemler, J. A., Dosey, A. M., Narayan, A. R., Hakansson, K., Sherman, D. H., Smith, J. L., and Skiniotis, G. Structural rearrangements of a polyketide synthase module during its catalytic cycle, *Nature 510*, 560-564 (2014).
61. Fenwick, M. K., and Ealick, S. E. Towards the structural characterization of the human methyltransferome, *Curr Opin Struct Biol 53*, 12-21 (2018).

62. Kozbial, P. Z., and Mushegian, A. R. Natural history of *S*-adenosylmethionine-binding proteins, *BMC Structural Biology 5*, 19 (2005).
63. Gana, R., Rao, S., Huang, H., Wu, C., and Vasudevan, S. Structural and functional studies of *S*-adenosyl-L-methionine binding proteins: a ligand-centric approach, *BMC Struct Biol 13*, 6 (2013).
64. Struck, A. W., Thompson, M. L., Wong, L. S., and Micklefield, J. S-adenosyl-methionine-dependent methyltransferases: highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications, *Chembiochem 13*, 2642-2655 (2012).
65. Liscombe, D. K., Louie, G. V., and Noel, J. P. Architectures, mechanisms and molecular evolution of natural product methyltransferases, *Nat Prod Rep 29*, 1238-1250 (2012).
66. Akey, D. L., Li, S., Konwerski, J. R., Confer, L. A., Bernard, S. M., Anzai, Y., Kato, F., Sherman, D. H., and Smith, J. L. A new structural form in the SAM/metal-dependent *O*-methyltransferase family: MycE from the mycinamicin biosynthetic pathway, *J Mol Biol 413*, 438-450 (2011).
67. Bernard, S. M., Akey, D. L., Tripathi, A., Park, S. R., Konwerski, J. R., Anzai, Y., Li, S., Kato, F., Sherman, D. H., and Smith, J. L. Structural basis of substrate specificity and regiochemistry in the MycF/TylF family of sugar *O*-methyltransferases, *ACS Chem Biol 10*, 1340-1351 (2015).
68. Singh, S., McCoy, J. G., Zhang, C., Bingman, C. A., Phillips, G. N., Jr., and Thorson, J. S. Structure and mechanism of the rebeccamycin sugar 4'-*O*-methyltransferase RebM, *J Biol Chem 283*, 22628-22636 (2008).
69. Jansson, A., Koskiniemi, H., Mantsala, P., Niemi, J., and Schneider, G. Crystal structure of a ternary complex of DnrK, a methyltransferase in daunorubicin biosynthesis, with bound products, *J Biol Chem 279*, 41149-41156 (2004).
70. Kopycki, J. G., Stubbs, M. T., Brandt, W., Hagemann, M., Porzel, A., Schmidt, J., Schliemann, W., Zenk, M. H., and Vogt, T. Functional and structural characterization of a cation-dependent *O*-methyltransferase from the cyanobacterium *Synechocystis* sp. strain PCC 6803, *J Biol Chem 283*, 20888-20896 (2008).
71. Pacholec, M., Tao, J., and Walsh, C. T. CouO and NovO: *C*-methyltransferases for tailoring the aminocoumarin scaffold in coumermycin and novobiocin antibiotic biosynthesis, *Biochemistry 44*, 14969-14976 (2005).
72. Koksal, M., Chou, W. K., Cane, D. E., and Christianson, D. W. Structure of geranyl diphosphate *C*-methyltransferase from *Streptomyces coelicolor* and implications for the mechanism of isoprenoid modification, *Biochemistry 51*, 3003-3010 (2012).
73. Stroupe, M. E., Leech, H. K., Daniels, D. S., Warren, M. J., and Getzoff, E. D. CysG structure reveals tetrapyrrole-binding features and novel regulation of siroheme biosynthesis, *Nat Struct Biol 10*, 1064-1073 (2003).
74. Zou, X. W., Liu, Y. C., Hsu, N. S., Huang, C. J., Lyu, S. Y., Chan, H. C., Chang, C. Y., Yeh, H. W., Lin, K. H., Wu, C. J., Tsai, M. D., and Li, T. L. Structure and mechanism of a nonhaem-iron SAM-dependent *C*-methyltransferase and its engineering to a hydratase and an *O*-methyltransferase, *Acta Crystallogr D Biol Crystallogr 70*, 1549-1560 (2014).
75. O'Brien, D. P., Kirkpatrick, P. N., O'Brien, S. W., Staroske, T., Richardson, T. I., Evans, D. A., Hopkinson, A., Spencer, J. B., and Williams, D. H. Expression and assay of an *N*-methyltransferase involved in the biosynthesis of a vancomycin group antibiotic, *Chem Commun (Camb)*, 103-104 (2000).

76. Lee, J., Hao, Y., Blair, P. M., Melby, J. O., Agarwal, V., Burkhart, B. J., Nair, S. K., and Mitchell, D. A. Structural and functional insight into an unexpectedly selective *N*-methyltransferase involved in plantazolicin biosynthesis, *Proc Natl Acad Sci U S A 110*, 12954-12959 (2013).
77. Ramm, S., Krawczyk, B., Muhlenweg, A., Poch, A., Mosker, E., and Sussmuth, R. D. A Self-Sacrificing N-Methyltransferase Is the Precursor of the Fungal Natural Product Omphalotin, *Angew Chem Int Ed Engl 56*, 9994-9997 (2017).
78. van der Velden, N. S., Kalin, N., Helf, M. J., Piel, J., Freeman, M. F., and Kunzler, M. Autocatalytic backbone *N*-methylation in a family of ribosomal peptide natural products, *Nat Chem Biol 13*, 833-835 (2017).
79. Hornbogen, T., Riechers, S. P., Prinz, B., Schultchen, J., Lang, C., Schmidt, S., Mugge, C., Turkanovic, S., Sussmuth, R. D., Tauberger, E., and Zocher, R. Functional characterization of the recombinant *N*-methyltransferase domain from the multienzyme enniatin synthetase, *Chembiochem 8*, 1048-1054 (2007).
80. Burgie, E. S., and Holden, H. M. Three-dimensional structure of DesVI from *Streptomyces venezuelae*: a sugar *N*,*N*-dimethyltransferase required for dTDP-desosamine biosynthesis, *Biochemistry 47*, 3982-3988 (2008).
81. Carney, A. E., and Holden, H. M. Molecular architecture of TylM1 from *Streptomyces fradiae*: an *N*,*N*-dimethyltransferase involved in the production of dTDP-D-mycaminose, *Biochemistry 50*, 780-787 (2011).
82. Hendrickson, L., Davis, C. R., Roach, C., Nguyen, D. K., Aldrich, T., McAda, P. C., and Reeves, C. D. Lovastatin biosynthesis in *Aspergillus terreus*: characterization of blocked mutants, enzyme activities and a multifunctional polyketide synthase gene, *Chem Biol 6*, 429-439 (1999).
83. Gehring, A. M., DeMoll, E., Fetherston, J. D., Mori, I., Mayhew, G. F., Blattner, F. R., Walsh, C. T., and Perry, R. D. Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*, *Chem Biol 5*, 573-586 (1998).
84. Winter, J. M., Sato, M., Sugimoto, S., Chiou, G., Garg, N. K., Tang, Y., and Watanabe, K. Identification and characterization of the chaetoviridin and chaetomugilin gene cluster in *Chaetomium globosum* reveal dual functions of an iterative highly-reducing polyketide synthase, *J Am Chem Soc 134*, 17900-17903 (2012).
85. Abe, Y., Suzuki, T., Ono, C., Iwamoto, K., Hosobuchi, M., and Yoshikawa, H. Molecular cloning and characterization of an ML-236B (compactin) biosynthetic gene cluster in *Penicillium citrinum*, *Mol Genet Genomics 267*, 636-646 (2002).
86. Proctor, R. H., Desjardins, A. E., Plattner, R. D., and Hohn, T. M. A polyketide synthase gene required for biosynthesis of fumonisin mycotoxins in *Gibberella fujikuroi* mating population A, *Fungal Genet Biol 27*, 100-112 (1999).
87. Eley, K. L., Halo, L. M., Song, Z., Powles, H., Cox, R. J., Bailey, A. M., Lazarus, C. M., and Simpson, T. J. Biosynthesis of the 2-pyridone tenellin in the insect pathogenic fungus Beauveria bassiana, *Chembiochem 8*, 289-297 (2007).
88. Cox, R. J., Glod, F., Hurley, D., Lazarus, C. M., Nicholson, T. P., Rudd, B. A., Simpson, T. J., Wilkinson, B., and Zhang, Y. Rapid cloning and expression of a fungal polyketide synthase gene involved in squalestatin biosynthesis, *Chem Commun (Camb)*, 2260-2261 (2004).
89. Magarvey, N. A., Beck, Z. Q., Golakoti, T., Ding, Y., Huber, U., Hemscheidt, T. K., Abelson, D., Moore, R. E., and Sherman, D. H. Biosynthetic characterization and

chemoenzymatic assembly of the cryptophycins. Potent anticancer agents from cyanobionts, *ACS Chem Biol 1*, 766-779 (2006).
90. Ramaswamy, A. V., Sorrels, C. M., and Gerwick, W. H. Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*, *J Nat Prod 70*, 1977-1986 (2007).
91. Tillett, D., Dittmann, E., Erhard, M., von Dohren, H., Borner, T., and Neilan, B. A. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system, *Chem Biol 7*, 753-764 (2000).
92. Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C., and Julien, B. Cloning and heterologous expression of the epothilone gene cluster, *Science 287*, 640-642 (2000).
93. Mattheus, W., Gao, L. J., Herdewijn, P., Landuyt, B., Verhaegen, J., Masschelein, J., Volckaert, G., and Lavigne, R. Isolation and purification of a new kalimantacin/batumin-related polyketide antibiotic and elucidation of its biosynthesis gene cluster, *Chem Biol 17*, 149-159 (2010).
94. Moebius, N., Ross, C., Scherlach, K., Rohm, B., Roth, M., and Hertweck, C. Biosynthesis of the respiratory toxin bongkrekic acid in the pathogenic bacterium *Burkholderia gladioli*, *Chem Biol 19*, 1164-1174 (2012).
95. Piel, J. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles, *Proc Natl Acad Sci U S A 99*, 14002-14007 (2002).
96. Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N., and Matsunaga, S. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*, *Proc Natl Acad Sci U S A 101*, 16222-16227 (2004).
97. Simunovic, V., Zapp, J., Rachid, S., Krug, D., Meiser, P., and Muller, R. Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and trans-acting acyltransferases, *Chembiochem 7*, 1206-1220 (2006).
98. Partida-Martinez, L. P., de Looss, C. F., Ishida, K., Ishida, M., Roth, M., Buder, K., and Hertweck, C. Rhizoxin, the first mycotoxin isolated from the zygomycota, is not a fungal metabolite but is produced by bacterial endosymbionts, *Appl Environ Microbiol 73*, 793-797 (2007).
99. Cacho, R. A., Thuss, J., Xu, W., Sanichar, R., Gao, Z., Nguyen, A., Vederas, J. C., and Tang, Y. Understanding programming of fungal iterative polyketide synthases: the biochemical basis for regioselectivity by the methyltransferase domain in the lovastatin megasynthase, *J Am Chem Soc 137*, 15688-15691 (2015).
100. Miller, D. A., Luo, L., Hillson, N., Keating, T. A., and Walsh, C. T. Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of *Yersinia pestis*, *Chem Biol 9*, 333-344 (2002).
101. Poust, S., Phelan, R. M., Deng, K., Katz, L., Petzold, C. J., and Keasling, J. D. Divergent mechanistic routes for the formation of gem-dimethyl groups in the biosynthesis of complex polyketides, *Angew Chem Int Ed Engl 54*, 2370-2373 (2015).
102. Stevens, D. C., Wagner, D. T., Manion, H. R., Alexander, B. K., and Keatinge-Clay, A. T. Methyltransferases excised from trans-AT polyketide synthases operate on *N*-acetylcysteamine-bound substrates, *J Antibiot (Tokyo)* (2016).
103. Wagner, D. T., Stevens, D. C., Mehaffey, M. R., Manion, H. R., Taylor, R. E., Brodbelt, J. S., and Keatinge-Clay, A. T. α-methylation follows condensation in the gephyronic acid modular polyketide synthase, *Chem Commun (Camb) 52*, 8822-8825 (2016).

104. Zimmermann, K., Engeser, M., Blunt, J. W., Munro, M. H., and Piel, J. Pederin-type pathways of uncultivated bacterial symbionts: analysis of *O*-methyltransferases and generation of a biosynthetic hybrid, *J Am Chem Soc 131*, 2780-2781 (2009).
105. Liu, Q., Yao, F., Chooi, Y. H., Kang, Q., Xu, W., Li, Y., Shao, Y., Shi, Y., Deng, Z., Tang, Y., and You, D. Elucidation of Piericidin A1 biosynthetic locus revealed a thioesterase-dependent mechanism of alpha-pyridone ring formation, *Chem Biol 19*, 243-253 (2012).
106. Schwecke, T., Aparicio, J. F., Molnar, I., Konig, A., Khaw, L. E., Haydock, S. F., Oliynyk, M., Caffrey, P., Cortes, J., Lester, J. B., and et al. The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin, *Proc Natl Acad Sci U S A 92*, 7839-7843 (1995).
107. Weinig, S., Hecht, H. J., Mahmud, T., and Muller, R. Melithiazol biosynthesis: further insights into myxobacterial PKS/NRPS systems and evidence for a new subclass of methyl transferases, *Chem Biol 10*, 939-952 (2003).
108. Feng, Z., Qi, J., Tsuge, T., Oba, Y., Kobayashi, T., Suzuki, Y., Sakagami, Y., and Ojika, M. Construction of a bacterial artificial chromosome library for a myxobacterium of the genus *Cystobacter* and characterization of an antibiotic biosynthetic gene cluster, *Biosci Biotechnol Biochem 69*, 1372-1380 (2005).
109. Buntin, K., Weissman, K. J., and Muller, R. An unusual thioesterase promotes isochromanone ring formation in ajudazol biosynthesis, *Chembiochem 11*, 1137-1146 (2010).
110. Muller, S., Rachid, S., Hoffmann, T., Surup, F., Volz, C., Zaburannyi, N., and Muller, R. Biosynthesis of crocacin involves an unusual hydrolytic release domain showing similarity to condensation domains, *Chem Biol 21*, 855-865 (2014).
111. Krastel, P., Roggo, S., Schirle, M., Ross, N. T., Perruccio, F., Aspesi, P., Jr., Aust, T., Buntin, K., Estoppey, D., Liechty, B., Mapa, F., Memmert, K., Miller, H., Pan, X., Riedl, R., Thibaut, C., Thomas, J., Wagner, T., Weber, E., Xie, X., Schmitt, E. K., and Hoepfner, D. Nannocystin A: an elongation factor 1 inhibitor from myxobacteria with differential anti-cancer properties, *Angew Chem Int Ed Engl 54*, 10149-10154 (2015).
112. Steinmetz, H., Li, J., Fu, C., Zaburannyi, N., Kunze, B., Harmrolfs, K., Schmitt, V., Herrmann, J., Reichenbach, H., Hofle, G., Kalesse, M., and Muller, R. Isolation, structure elucidation, and (bio)synthesis of haprolid, a cell-type-specific myxobacterial cytotoxin, *Angew Chem Int Ed Engl 55*, 10113-10117 (2016).
113. Silakowski, B., Schairer, H. U., Ehret, H., Kunze, B., Weinig, S., Nordsiek, G., Brandt, P., Blocker, H., Hofle, G., Beyer, S., and Muller, R. New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1, *J Biol Chem 274*, 37391-37399 (1999).
114. Kwan, J. C., Donia, M. S., Han, A. W., Hirose, E., Haygood, M. G., and Schmidt, E. W. Genome streamlining and chemical defense in a coral reef symbiosis, *Proc Natl Acad Sci U S A 109*, 20655-20660 (2012).
115. Wakimoto, T., Egami, Y., Nakashima, Y., Wakimoto, Y., Mori, T., Awakawa, T., Ito, T., Kenmoku, H., Asakawa, Y., Piel, J., and Abe, I. Calyculin biogenesis from a pyrophosphate protoxin produced by a sponge symbiont, *Nat Chem Biol 10*, 648-655 (2014).
116. Kampa, A., Gagunashvili, A. N., Gulder, T. A., Morinaka, B. I., Daolio, C., Godejohann, M., Miao, V. P., Piel, J., and Andresson, O. Metagenomic natural product discovery in

lichen provides evidence for a family of biosynthetic pathways in diverse symbioses, *Proc Natl Acad Sci U S A 110*, E3129-3137 (2013).
117. Skiba, M. A., Sikkema, A. P., Fiers, W. D., Gerwick, W. H., Sherman, D. H., Aldrich, C. C., and Smith, J. L. Domain organization and active site architecture of a polyketide synthase *C*-methyltransferase, *ACS Chem Biol 11*, 3319-3327 (2016).
118. Cai, W., and Zhang, W. Engineering modular polyketide synthases for production of biofuels and industrial chemicals, *Curr Opin Biotechnol 50*, 32-38 (2018).
119. Skiba, M. A., Sikkema, A. P., Moss, N. A., Tran, C. L., Sturgis, R. M., Gerwick, L., Gerwick, W. H., Sherman, D. H., and Smith, J. L. A mononuclear iron-dependent methyltransferase catalyzes initial steps in assembly of the apratoxin A polyketide starter unit, *ACS Chem Biol 12*, 3039-3048 (2017).
120. Skiba, M. A., Sikkema, A. P., Moss, N. A., Lowell, A. N., Su, M., Sturgis, R. M., Gerwick, L., Gerwick, W. H., Sherman, D. H., and Smith, J. L. Biosynthesis of *t*-Butyl in apratoxin A: functional analysis and architecture of a PKS loading module, *ACS Chem Biol 13*, 1640-1650 (2018).
121. Cox, R. J. Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets, *Org Biomol Chem 5*, 2010-2026 (2007).
122. Crawford, J. M., and Townsend, C. A. New insights into the formation of fungal aromatic polyketides, *Nat Rev Microbiol 8*, 879-889 (2010).
123. Kennedy, J., Auclair, K., Kendrew, S. G., Park, C., Vederas, J. C., and Hutchinson, C. R. Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis, *Science 284*, 1368-1372 (1999).
124. Ansari, M. Z., Sharma, J., Gokhale, R. S., and Mohanty, D. In silico analysis of methyltransferase domains involved in biosynthesis of secondary metabolites, *BMC Bioinformatics 9*, 454 (2008).
125. Kozbial, P. Z., and Mushegian, A. R. Natural history of *S*-adenosylmethionine-binding proteins, *BMC Struct Biol 5*, 19 (2005).
126. Gerwick, W. H., Proteau, P. J., Nagle, D. G., Hamel, E., Blokin, A., and Slate, D. L. Structure of curacin A, a novel antimitotic, antiproliferative, and brine shrimp toxic natural product from the marine cyanobacterium *Lyngbya majuscula*, *J Org Chem 59*, 1243-1245 (1994).
127. Blokhin, A. V., Yoo, H., Geralds, R. S., Nagle, D. G., Gerwick, W. H., and Hamel, E. Characterization of the interaction of the marine cyanobacterial natural prouduct curacin A with the colchicine site of tubulin and initial structure-activity studies with analogues, *Mol Pharmacol* (1995).
128. Kabsch, W. Xds, *Acta Crystallogr D Biol Crystallogr 66*, 125-132 (2010).
129. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification, *Acta Crystallogr D Biol Crystallogr 66*, 479-485 (2010).
130. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr 66*, 213-221 (2010).
131. Terwilliger, T. C. SOLVE and RESOLVE: automated structure solution and density modification, *Methods Enzymol 374*, 22-37 (2003).

132. Emsley, P., and Cowtan, K. Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr 60*, 2126-2132 (2004).
133. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr D Biol Crystallogr 66*, 12-21 (2010).
134. Trott, O., and Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J Comput Chem 31*, 455-461 (2010).
135. Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., Gora, A., Sustr, V., Klvana, M., Medek, P., Biedermannova, L., Sochor, J., and Damborsky, J. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures, *PLoS Comput Biol 8*, e1002708 (2012).
136. Schrodinger, LLC. (2015) The PyMOL Molecular Graphics System, Version 1.8.
137. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. Clustal W and Clustal X version 2.0, *Bioinformatics 23*, 2947-2948 (2007).
138. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench, *Bioinformatics 25*, 1189-1191 (2009).
139. Roy, A., Kucukural, A., and Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction, *Nat Protoc 5*, 725-738 (2010).
140. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. The I-TASSER Suite: protein structure and function prediction, *Nat Methods 12*, 7-8 (2015).
141. Zhang, Y. I-TASSER server for protein 3D structure prediction, *BMC Bioinformatics 9*, 40 (2008).
142. Sánchez, C., Du, L., Edwards, D. J., Toney, M. D., and Shen, B. Cloning and characterization of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin, *Chem Biol 8*, 725-738 (2001).
143. Li, Y., Dodge, G. J., Fiers, W. D., Fecik, R. A., Smith, J. L., and Aldrich, C. C. Functional characterization of a dehydratase domain from the pikromycin polyketide synthase, *J Am Chem Soc 137*, 7003-7006 (2015).
144. Piasecki, S. K., Zheng, J., Axelrod, A. J., Detelich, M. E., and Keatinge-Clay, A. T. Structural and functional studies of a trans-acyltransferase polyketide assembly line enzyme that catalyzes stereoselective alpha- and beta-ketoreduction, *Proteins 82*, 2067-2077 (2014).
145. Zeng, J., Wagner, D. T., Zhang, Z., Moretto, L., Addison, J. D., and Keatinge-Clay, A. T. Portability and structure of the four-helix bundle docking domains of trans-acyltransferase modular polyketide synthases, *ACS Chem Biol* (2016).
146. Zheng, J., Taylor, C. A., Piasecki, S. K., and Keatinge-Clay, A. T. Structural and functional analysis of A-type ketoreductases from the amphotericin modular polyketide synthase, *Structure 18*, 913-922 (2010).

147. Zheng, J., Piasecki, S. K., and Keatinge-Clay, A. T. Structural studies of an A2-type modular polyketide synthase ketoreductase reveal features controlling alpha-substituent stereochemistry, *ACS Chem Biol 8*, 1964-1971 (2013).
148. Horowitz, S., Dirk, L. M., Yesselman, J. D., Nimtz, J. S., Adhikari, U., Mehl, R. A., Scheiner, S., Houtz, R. L., Al-Hashimi, H. M., and Trievel, R. C. Conservation and functional importance of carbon-oxygen hydrogen bonding in AdoMet-dependent methyltransferases, *J Am Chem Soc 135*, 15536-15548 (2013).
149. Huang, C. C., Smith, C. V., Glickman, M. S., Jacobs, W. R., Jr., and Sacchettini, J. C. Crystal structures of mycolic acid cyclopropane synthases from *Mycobacterium tuberculosis*, *J Biol Chem 277*, 11559-11569 (2002).
150. Meluzzi, D., Zheng, W. H., Hensler, M., Nizet, V., and Dorrestein, P. C. Top-down mass spectrometry on low-resolution instruments: characterization of phosphopantetheinylated carrier domains in polyketide and non-ribosomal biosynthetic pathways, *Bioorg Med Chem Lett 18*, 3107-3111 (2008).
151. Ikeda, H., Nonomiya, T., Usami, M., Ohta, T., and Omura, S. Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*, *Proc Natl Acad Sci U S A 96*, 9509-9514 (1999).
152. Grimm, A., Madduri, K., Ali, A., and Hutchinson, C. R. Characterization of the *Streptomyces peucetius* ATCC 29050 genes encoding doxorubicin polyketide synthase, *Gene 151*, 1-10 (1994).
153. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J., and Katz, L. Modular organization of genes required for complex polyketide biosynthesis, *Science 252*, 675-679 (1991).
154. Keatinge-Clay, A. T. Stereocontrol within polyketide assembly lines, *Nat Prod Rep 33*, 141-149 (2016).
155. Xie, X., Khosla, C., and Cane, D. E. Elucidation of the stereospecificity of *C*-methyltransferases from trans-AT polyketide synthases, *J Am Chem Soc 139*, 6102-6105 (2017).
156. Zheng, J., and Keatinge-Clay, A. T. The status of type I polyketide synthase ketoreductases, *Med. Chem. Commum. 4*, 34-40 (2013).
157. Caffrey, P. Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases, *Chembiochem 4*, 654-657 (2003).
158. Valenzano, C. R., You, Y. O., Garg, A., Keatinge-Clay, A., Khosla, C., and Cane, D. E. Stereospecificity of the dehydratase domain of the erythromycin polyketide synthase, *J Am Chem Soc 132*, 14697-14699 (2010).
159. Gay, D., You, Y. O., Keatinge-Clay, A., and Cane, D. E. Structure and stereospecificity of the dehydratase domain from the terminal module of the rifamycin polyketide synthase, *Biochemistry 52*, 8916-8928 (2013).
160. Fiers, W. D., Dodge, G. J., Sherman, D. H., Smith, J. L., and Aldrich, C. C. Vinylogous dehydration by a polyketide dehydratase domain in curacin biosynthesis, *J Am Chem Soc 138*, 16024-16036 (2016).
161. Xie, X., and Cane, D. E. Stereospecific formation of Z-trisubstituted double bonds by the successive action of ketoreductase and dehydratase domains from trans-AT polyketide synthases, *Biochemistry 57*, 3126-3129 (2018).

162. Zhang, L., Ji, J., Yuan, M., Feng, Y., Wang, L., Deng, Z., Bai, L., and Zheng, J. Stereospecificity of enoylreductase domains from modular polyketide synthases, *ACS Chem Biol 13*, 871-875 (2018).
163. Kwan, D. H., Sun, Y., Schulz, F., Hong, H., Popovic, B., Sim-Stark, J. C., Haydock, S. F., and Leadlay, P. F. Prediction and manipulation of the stereochemistry of enoylreduction in modular polyketide synthases, *Chem Biol 15*, 1231-1240 (2008).
164. Klaus, M., and Grininger, M. Engineering strategies for rational polyketide synthase design, *Nat Prod Rep* (2018).
165. Storm, P. A., Herbst, D. A., Maier, T., and Townsend, C. A. Functional and structural analysis of programmed *C*-methylation in the biosynthesis of the fungal polyketide citrinin, *Cell Chem Biol 24*, 316-325 (2017).
166. Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F. R., and Donnelly, M. I. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site, *Protein Expr Purif 25*, 8-15 (2002).
167. DelProposto, J., Majmudar, C. Y., Smith, J. L., and Brown, W. C. Mocr: a novel fusion tag for enhancing solubility that is compatible with structural biology applications, *Protein Expr Purif 63*, 40-49 (2009).
168. Skiba, M. A., Maloney, F. P., Dan, Q., Fraley, A. E., Aldrich, C. C., Smith, J. L., and Brown, W. C. PKS-NRPS enzymology and structural biology: considerations in protein production, *Methods Enzymol 604*, 45-88 (2018).
169. Dorrestein, P. C., Bumpus, S. B., Calderone, C. T., Garneau-Tsodikova, S., Aron, Z. D., Straight, P. D., Kolter, R., Walsh, C. T., and Kelleher, N. L. Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry, *Biochemistry 45*, 12756-12766 (2006).
170. Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., Terwilliger, T. C., and Adams, P. D. Automated structure solution with the PHENIX suite, *Methods Mol Biol 426*, 419-435 (2008).
171. Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard, *Acta Crystallogr D Biol Crystallogr 64*, 61-69 (2008).
172. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. Phaser crystallographic software, *J Appl Crystallogr 40*, 658-674 (2007).
173. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. Towards automated crystallographic structure refinement with phenix.refine, *Acta Crystallogr D Biol Crystallogr 68*, 352-367 (2012).
174. Holm, L., and Rosenstrom, P. Dali server: conservation mapping in 3D, *Nucleic Acids Res 38*, W545-549 (2010).
175. Lee, S. G., Kim, Y., Alpert, T. D., Nagata, A., and Jez, J. M. Structure and reaction mechanism of phosphoethanolamine methyltransferase from the malaria parasite *Plasmodium falciparum*: an antiparasitic drug target, *J Biol Chem 287*, 1426-1434 (2012).

176. Fage, C. D., Isiorho, E. A., Liu, Y., Wagner, D. T., Liu, H. W., and Keatinge-Clay, A. T. The structure of SpnF, a standalone enzyme that catalyzes [4 + 2] cycloaddition, *Nat Chem Biol 11*, 256-258 (2015).
177. Donadio, S., and Katz, L. Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*, *Gene 111*, 51-60 (1992).
178. Moore, B. S., and Hertweck, C. Biosynthesis and attachment of novel bacterial polyketide synthase starter units, *Nat Prod Rep 19*, 70-99 (2002).
179. Cane, D. E., Liang, T. C., Kaplan, L., Nallin, M. K., Schulman, M. D., Hensens, O. D., Douglas, A. W., and G, A.-S. Biosynthetic origin of the carbon skeleton and oxygen atoms of the avermectins, *J Am Chem Soc 105*, 4110-4112 (1983).
180. Hafner, E. W., Holley, B. W., Holdom, K. S., Lee, S. E., Wax, R. G., Beck, D., McArthur, H. A., and Wernau, W. C. Branched-chain fatty acid requirement for avermectin production by a mutant of *Streptomyces avermitilis* lacking branched-chain 2-oxo acid dehydrogenase activity, *J Antibiot (Tokyo) 44*, 349-356 (1991).
181. Pulsawat, N., Kitani, S., Kinoshita, H., Lee, C. K., and Nihira, T. Identification of the bkdAB gene cluster, a plausible source of the starter-unit for virginiamycin M production in *Streptomyces virginiae*, *Arch Microbiol 187*, 459-466 (2007).
182. Pulsawat, N., Kitani, S., and Nihira, T. Characterization of biosynthetic gene cluster for the production of virginiamycin M, a streptogramin type A antibiotic, in *Streptomyces virginiae*, *Gene 393*, 31-42 (2007).
183. Froese, D. S., Forouhar, F., Tran, T. H., Vollmar, M., Kim, Y. S., Lew, S., Neely, H., Seetharaman, J., Shen, Y., Xiao, R., Acton, T. B., Everett, J. K., Cannone, G., Puranik, S., Savitsky, P., Krojer, T., Pilka, E. S., Kiyani, W., Lee, W. H., Marsden, B. D., von Delft, F., Allerston, C. K., Spagnolo, L., Gileadi, O., Montelione, G. T., Oppermann, U., Yue, W. W., and Tong, L. Crystal structures of malonyl-coenzyme A decarboxylase provide insights into its catalytic mechanism and disease-causing mutations, *Structure 21*, 1182-1192 (2013).
184. Leao, T., Castelao, G., Korobeynikov, A., Monroe, E. A., Podell, S., Glukhov, E., Allen, E. E., Gerwick, W. H., and Gerwick, L. Comparative genomics uncovers the prolific and distinctive metabolic potential of the cyanobacterial genus *Moorea*, *Proc Natl Acad Sci U S A 114*, 3198-3203 (2017).
185. Pfeifer, B. A., and Khosla, C. Biosynthesis of polyketides in heterologous hosts, *Microbiol Mol Biol Rev 65*, 106-118 (2001).
186. Dayem, L. C., Carney, J. R., Santi, D. V., Pfeifer, B. A., Khosla, C., and Kealey, J. T. Metabolic engineering of a methylmalonyl-CoA mutase-epimerase pathway for complex polyketide biosynthesis in *Escherichia coli*, *Biochemistry 41*, 5193-5201 (2002).
187. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome, *Proc Natl Acad Sci U S A 98*, 10037-10041 (2001).
188. Kraft, A. S., Smith, J. B., and Berkow, R. L. Bryostatin, an activator of the calcium phospholipid-dependent protein kinase, blocks phorbol ester-induced differentiation of human promyelocytic leukemia cells HL-60, *Proc Natl Acad Sci U S A 83*, 1334-1338 (1986).
189. Vidgren, J., Svensson, L. A., and Liljas, A. Crystal structure of catechol *O*-methyltransferase, *Nature 368*, 354-358 (1994).

190. Armstrong, R. N. Mechanistic diversity in a metalloenzyme superfamily, *Biochemistry 39*, 13625-13632 (2000).
191. Hashimoto, K., Suzuki, H., Taniguchi, K., Noguchi, T., Yohda, M., and Odaka, M. Catalytic mechanism of nitrile hydratase proposed by time-resolved X-ray crystallography using a novel substrate, tert-butylisonitrile, *J Biol Chem 283*, 36617-36623 (2008).
192. Nagashima, S., Nakasako, M., Dohmae, N., Tsujimura, M., Takio, K., Odaka, M., Yohda, M., Kamiya, N., and Endo, I. Novel non-heme iron center of nitrile hydratase with a claw setting of oxygen atoms, *Nat Struct Biol 5*, 347-351 (1998).
193. Arnett, E. M., Maroldo, S. G., Schilling, S. L., and Harrelson, J. A. Ion pairing and reactivity of enolate anions. 5. Thermodynamics of ionization of β-di- and tricarbonyl compounds in dimethyl sulfoxide solution and ion pairing of their alkali salts, *J Am Chem Soc 106*, 6759-6767 (1984).
194. Mazur, M. T., Walsh, C. T., and Kelleher, N. L. Site-specific observation of acyl intermediate processing in thiotemplate biosynthesis by fourier transform mass spectrometry: the polyketide module of yersiniabactin synthetase, *Biochemistry 42*, 13393-13400 (2003).
195. Carreras, C. W., and Khosla, C. Purification and in vitro reconstitution of the essential protein components of an aromatic polyketide synthase, *Biochemistry 37*, 2084-2088 (1998).
196. Florova, G., Kazanina, G., and Reynolds, K. A. Enzymes involved in fatty acid and polyketide biosynthesis in *Streptomyces glaucescens*: role of FabH and FabD and their acyl carrier protein specificity, *Biochemistry 41*, 10462-10471 (2002).
197. Ishikawa, F., Sugimoto, H., and Kakeya, H. In vitro investigation of crosstalk between fatty acid and polyketide synthases in the andrimid biosynthetic assembly line, *Chembiochem 17*, 2137-2142 (2016).
198. Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H. G., and Prinsep, M. R. Marine natural products, *Nat Prod Rep 34*, 235-294 (2017).
199. Gerwick, W. H., and Moore, B. S. Lessons from the past and charting the future of marine natural products drug discovery and chemical biology, *Chem Biol 19*, 85-98 (2012).
200. Luesch, H., Yoshida, W. Y., Moore, R. E., Paul, V. J., and Corbett, T. H. Total structure determination of apratoxin A, a potent novel cytotoxin from the marine cyanobacterium *Lyngbya majuscula*, *J Am Chem Soc 123*, 5418-5423 (2001).
201. Crews, P., Farias, J. J., Emrich, R., and Keifer, P. A. Milnamide A, an unusual cytotoxic tripeptide from the marine sponge *Auletta cf. constricta*, *J Org Chem 59*, 2932-2934 (1994).
202. Parent, A., Guillot, A., Benjdia, A., Chartier, G., Leprince, J., and Berteau, O. The B12-radical SAM enzyme PoyC catalyzes valine Cβ-methylation during polytheonamide biosynthesis, *J Am Chem Soc 138*, 15515-15518 (2016).
203. Freeman, M. F., Helf, M. J., Bhushan, A., Morinaka, B. I., and Piel, J. Seven enzymes create extraordinary molecular complexity in an uncultivated bacterium, *Nat Chem 9*, 387-395 (2017).
204. Harunari, E., Komaki, H., and Igarashi, Y. Biosynthetic origin of butyrolactol A, an antifungal polyketide produced by a marine-derived *Streptomyces*, *Beilstein J Org Chem 13*, 441-450 (2017).

205. Ishibashi, M., Yamagishi, E., and Kobayashi, J. Topsentinols A-J, new sterols with highly branched side chains from marine sponge *Topsentia sp.*, *Chem Pharm Bull 45*, 1435-1438 (1997).
206. Mori, S., Williams, H., Cagle, D., Karanovich, K., Horgen, F. D., Smith, R., III, and Watanabe, C. M. Macrolactone nuiapolide, isolated from a Hawaiian marine cyanobacterium, exhibits anti-chemotactic activity, *Mar Drugs 13*, 6274-6290 (2015).
207. Ogawa, H., Iwasaki, A., Sumimoto, S., Kanamori, Y., Ohno, O., Iwatsuki, M., Ishiyama, A., Hokari, R., Otoguro, K., Omura, S., and Suenaga, K. Janadolide, a cyclic polyketide-peptide hybrid possessing a tert-butyl group from an *Okeania sp.* marine cyanobacterium, *J Nat Prod 79*, 1862-1866 (2016).
208. Pettit, G. R., Gao, F., Sengupta, D., Coll, J. C., Herald, C. L., Doubek, D. L., Schmidt, J. M., Van Camp, J. R., Rudloe, J. J., and Nieman, R. A. Isolation and structure of bryostatins 14 and 15, *Tetrahedron 47*, 3601-3610 (1991).
209. Klein, D., Braekman, J. C., Daloze, D., Hoffmann, L., Castillo, G., and Demoulin, V. V. Madangolide and laingolide A, two novel macrolides from *Lyngbya bouillonii J Nat Prod 62*, 934-936 (1999).
210. Salvador-Reyes, L. A., Sneed, J., Paul, V. J., and Luesch, H. Amantelides A and B, polyhydroxylated macrolides with differential broad-spectrum cytotoxicity from a guamanian marine cyanobacterium, *J Nat Prod 78*, 1957-1962 (2015).
211. Shao, C. L., Linington, R. G., Balunas, M. J., Centeno, A., Boudreau, P., Zhang, C., Engene, N., Spadafora, C., Mutka, T. S., Kyle, D. E., Gerwick, L., Wang, C. Y., and Gerwick, W. H. Bastimolide A, a potent antimalarial polyhydroxy macrolide from the marine cyanobacterium *Okeania hirsuta*, *J Org Chem 80*, 7849-7855 (2015).
212. Pereira, A. R., Cao, Z., Engene, N., Soria-Mercado, I. E., Murray, T. F., and Gerwick, W. H. Palmyrolide A, an unusually stabilized neuroactive macrolide from Palmyra Atoll cyanobacteria, *Org Lett 12*, 4490-4493 (2010).
213. Orjala, J., Nagle, D. G., Hsu, V. L., and Gerwick, W. H. Antillatoxin: an exceptionally ichthyotoxic cyclic lipopeptide from the tropical cyanobacterium *Lyngbya majuscula*, *J Am Chem Soc 117*, 8281-8282 (1995).
214. Nagle, D. G., Paul, V. J., and Roberts, M. A. Ypaoamide, a new broadly acting feeding deterrent from the marine cyanobacterium *Lyngbya majuscula*, *Tetrahedron Letters 37*, 6263-6266 (1996).
215. Gallimore, W. A., Galario, D. L., Lacy, C., Zhu, Y., and Scheuer, P. J. Two complex proline esters from the sea hare *Stylocheilus longicauda*, *J Nat Prod 63*, 1022-1026 (2000).
216. Teruya, T., Sasaki, H., Fukazawa, H., and Suenaga, K. Bisebromoamide, a potent cytotoxic peptide from the marine cyanobacterium *Lyngbya sp.*: isolation, stereostructure, and biological activity, *Org Lett 11*, 5062-5065 (2009).
217. Bisel, P., Al-Momani, L., and Muller, M. The tert-butyl group in chemistry and biology, *Org Biomol Chem 6*, 2655-2665 (2008).
218. Nakanishi, K. The ginkgolides, *Pure Appl Chem 14*, 89-113 (1967).
219. Chun, S. W., Hinze, M. E., Skiba, M. A., and Narayan, A. R. H. Chemistry of a unique polyketide-like synthase, *J Am Chem Soc 140*, 2430-2433 (2018).
220. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., and Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*, *Science 291*, 1790-1792 (2001).

221. Bunkoczi, G., and Read, R. J. Improvement of molecular-replacement models with Sculptor, *Acta Crystallogr D Biol Crystallogr 67*, 303-312 (2011).
222. Ohi, M., Li, Y., Cheng, Y., and Walz, T. Negative staining and image classification - powerful tools in modern electron microscopy, *Biol Proced Online 6*, 23-34 (2004).
223. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination, *J Struct Biol 180*, 519-530 (2012).
224. Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. EMAN2: an extensible image processing suite for electron microscopy, *J Struct Biol 157*, 38-46 (2007).
225. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem 25*, 1605-1612 (2004).
226. Liutkeviciute, Z., Kriukiene, E., Licyte, J., Rudyte, M., Urbanaviciute, G., and Klimasauskas, S. Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases, *J Am Chem Soc 136*, 5884-5887 (2014).
227. Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W., and Roberts, R. J. Crystal structure of the HhaI DNA methyltransferase complexed with *S*-adenosyl-L-methionine, *Cell 74*, 299-307 (1993).
228. Keller, J. P., Smith, P. M., Benach, J., Christendat, D., deTitta, G. T., and Hunt, J. F. The crystal structure of MT0146/CbiT suggests that the putative precorrin-8w decarboxylase is a methyltransferase, *Structure 10*, 1475-1487 (2002).
229. Deery, E., Schroeder, S., Lawrence, A. D., Taylor, S. L., Seyedarabi, A., Waterman, J., Wilson, K. S., Brown, D., Geeves, M. A., Howard, M. J., Pickersgill, R. W., and Warren, M. J. An enzyme-trap approach allows isolation of intermediates in cobalamin biosynthesis, *Nat Chem Biol 8*, 933-940 (2012).
230. Luesch, H., Yoshida, W. Y., Moore, R. E., and Paul, V. J. New apratoxins of marine cyanobacterial origin from Guam and Palau, *Bioorg Med Chem 10*, 1973-1978 (2002).
231. Fick, R. J., Kroner, G. M., Nepal, B., Magnani, R., Horowitz, S., Houtz, R. L., Scheiner, S., and Trievel, R. C. Sulfur-oxygen chalcogen bonding mediates AdoMet recognition in the lysine methyltransferase SET7/9, *ACS Chem Biol 11*, 748-754 (2016).
232. Wesener, S. R., Potharla, V. Y., and Cheng, Y. Q. Reconstitution of the FK228 biosynthetic pathway reveals cross talk between modular polyketide synthases and fatty acid synthase, *Appl Environ Microbiol 77*, 1501-1507 (2011).
233. Kumar, P., Koppisch, A. T., Cane, D. E., and Khosla, C. Enhancing the modularity of the modular polyketide synthases: transacylation in modular polyketide synthases catalyzed by malonyl-CoA:ACP transacylase, *J Am Chem Soc 125*, 14307-14312 (2003).
234. Neuwald, A. F., and Landsman, D. GCN5-related histone *N*-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein, *Trends Biochem Sci 22*, 154-155 (1997).
235. Couture, J. F., and Trievel, R. C. Histone-modifying enzymes: encrypting an enigmatic epigenetic code, *Curr Opin Struct Biol 16*, 753-760 (2006).
236. Farazi, T. A., Waksman, G., and Gordon, J. I. Structures of *Saccharomyces cerevisiae N*-myristoyltransferase with bound myristoylCoA and peptide provide insights about substrate recognition and catalysis, *Biochemistry 40*, 6335-6343 (2001).

237. Benson, T. E., Prince, D. B., Mutchler, V. T., Curry, K. A., Ho, A. M., Sarver, R. W., Hagadorn, J. C., Choi, G. H., and Garlick, R. L. X-ray crystal structure of *Staphylococcus aureus* FemA, *Structure 10*, 1107-1115 (2002).
238. Van Wagoner, R. M., and Clardy, J. FeeM, an *N*-acyl amino acid synthase from an uncultured soil microbe: structure, mechanism, and acyl carrier protein binding, *Structure 14*, 1425-1435 (2006).
239. Watson, W. T., Minogue, T. D., Val, D. L., von Bodman, S. B., and Churchill, M. E. Structural basis and specificity of acyl-homoserine lactone signal production in bacterial quorum sensing, *Mol Cell 9*, 685-694 (2002).
240. Gould, T. A., Schweizer, H. P., and Churchill, M. E. Structure of the *Pseudomonas aeruginosa* acyl-homoserinelactone synthase LasI, *Mol Microbiol 53*, 1135-1146 (2004).
241. Hayaishi, O. Enzymatic decarboxylation of malonic acid, *J Biol Chem 215*, 125-136 (1955).
242. Aparicio, D., Perez-Luque, R., Carpena, X., Diaz, M., Ferrer, J. C., Loewen, P. C., and Fita, I. Structural asymmetry and disulfide bridges among subunits modulate the activity of human malonyl-CoA decarboxylase, *J Biol Chem 288*, 11907-11919 (2013).
243. An, J. H., and Kim, Y. S. A gene cluster encoding malonyl-CoA decarboxylase (MatA), malonyl-CoA synthetase (MatB) and a putative dicarboxylate carrier protein (MatC) in *Rhizobium trifolii*--cloning, sequencing, and expression of the enzymes in *Escherichia coli*, *Eur J Biochem 257*, 395-402 (1998).
244. Kim, Y. S., and Kolattukudy, P. E. Malonyl-CoA decarboxylase from the uropygial gland of waterfowl: purification, properties, immunological comparison, and role in regulating the synthesis of multimethyl-branched fatty acids, *Arch Biochem Biophys 190*, 585-597 (1978).
245. Sacksteder, K. A., Morrell, J. C., Wanders, R. J., Matalon, R., and Gould, S. J. MCD encodes peroxisomal and cytoplasmic forms of malonyl-CoA decarboxylase and is mutated in malonyl-CoA decarboxylase deficiency, *J Biol Chem 274*, 24461-24468 (1999).
246. Lombard, J., and Moreira, D. Early evolution of the biotin-dependent carboxylase family, *BMC Evol Biol 11*, 232 (2011).
247. Vetting, M. W., LP, S. d. C., Yu, M., Hegde, S. S., Magnet, S., Roderick, S. L., and Blanchard, J. S. Structure and functions of the GNAT superfamily of acetyltransferases, *Arch Biochem Biophys 433*, 212-226 (2005).
248. Oliynyk, M., Samborskyy, M., Lester, J. B., Mironenko, T., Scott, N., Dickens, S., Haydock, S. F., and Leadlay, P. F. Complete genome sequence of the erythromycin-producing bacterium Saccharopolyspora erythraea NRRL23338, *Nat Biotechnol 25*, 447-453 (2007).
249. Lazos, O., Tosin, M., Slusarczyk, A. L., Boakes, S., Cortes, J., Sidebottom, P. J., and Leadlay, P. F. Biosynthesis of the putative siderophore erythrochelin requires unprecedented crosstalk between separate nonribosomal peptide gene clusters, *Chem Biol 17*, 160-173 (2010).
250. Robbel, L., Helmetag, V., Knappe, T. A., and Marahiel, M. A. Consecutive enzymatic modification of ornithine generates the hydroxamate moieties of the siderophore erythrochelin, *Biochemistry 50*, 6073-6080 (2011).
251. Hsieh, Y. J., and Kolattukudy, P. E. Inhibition of erythromycin synthesis by disruption of malonyl-coenzyme A decarboxylase gene *eryM* in *Saccharopolyspora erythraea*, *J Bacteriol 176*, 714-724 (1994).

252. Card, G. L., Peterson, N. A., Smith, C. A., Rupp, B., Schick, B. M., and Baker, E. N. The crystal structure of Rv1347c, a putative antibiotic resistance protein from *Mycobacterium tuberculosis*, reveals a GCN5-related fold and suggests an alternative function in siderophore biosynthesis, *J Biol Chem 280*, 13978-13986 (2005).
253. Frankel, B. A., and Blanchard, J. S. Mechanistic analysis of *Mycobacterium tuberculosis* Rv1347c, a lysine N^{ε}-acyltransferase involved in mycobactin biosynthesis, *Arch Biochem Biophys 477*, 259-266 (2008).
254. Krithika, R., Marathe, U., Saxena, P., Ansari, M. Z., Mohanty, D., and Gokhale, R. S. A genetic locus required for iron acquisition in *Mycobacterium tuberculosis*, *Proc Natl Acad Sci U S A 103*, 2069-2074 (2006).
255. Weissman, K. J., Bycroft, M., Staunton, J., and Leadlay, P. F. Origin of starter units for erythromycin biosynthesis, *Biochemistry 37*, 11012-11017 (1998).
256. Tully, B. J., Wheat, C. G., Glazer, B. T., and Huber, J. A. A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer, *ISME J 12*, 1-16 (2018).
257. Chung, E. J., Park, J. A., Jeon, C. O., and Chung, Y. R. *Gynuella sunshinyii* gen. nov., sp. nov., an antifungal rhizobacterium isolated from a halophyte, *Carex scabrifolia* Steud, *Int J Syst Evol Microbiol 65*, 1038-1043 (2015).
258. Crits-Christoph, A., Diamond, S., Butterfield, C. N., Thomas, B. C., and Banfield, J. F. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis, *Nature 558*, 440-444 (2018).
259. Ling, S. K., Xia, J., Liu, Y., Chen, G. J., and Du, Z. J. *Agarilytica rhodophyticola* gen. nov., sp. nov., isolated from *Gracilaria blodgettii*, *Int J Syst Evol Microbiol 67*, 3778-3783 (2017).
260. Skyrud, W., Liu, J., Thankachan, D., Cabrera, M., Seipke, R. F., and Zhang, W. Biosynthesis of the 15-membered ring depsipeptide neoantimycin, *ACS Chem Biol 13*, 1398-1406 (2018).
261. Jackson, D. R., Shakya, G., Patel, A. B., Mohammed, L. Y., Vasilakis, K., Wattana-Amorn, P., Valentic, T. R., Milligan, J. C., Crump, M. P., Crosby, J., and Tsai, S. C. Structural and functional studies of the daunorubicin priming ketosynthase DpsC, *ACS Chem Biol 13*, 141-151 (2018).
262. Zhang, J., and Zheng, Y. G. SAM/SAH analogs as versatile tools for SAM-dependent methyltransferases, *ACS Chem Biol 11*, 583-597 (2016).
263. Cheng, Y. Single-particle cryo-EM at crystallographic resolution, *Cell 161*, 450-457 (2015).
264. McMahon, C., Baier, A. S., Pascolutti, R., Wegrecki, M., Zheng, S., Ong, J. X., Erlandson, S. C., Hilger, D., Rasmussen, S. G. F., Ring, A. M., Manglik, A., and Kruse, A. C. Yeast surface display platform for rapid discovery of conformationally selective nanobodies, *Nat Struct Mol Biol 25*, 289-296 (2018).